DOCUMENT RESUME

ED 215 010	UD 022 052
AUTHOR	Baker, Keith A.; de Kanter, Adriana A.
TITLE	Effectiveness of Bilingual Education: A Review of the Literature. Final Draft Report.
INSTITUTION	Department of Education, Washington, D.C. Office of Planning, Budget, and Evaluation.
PUB DATE	25 Sep 81
NOTE	213p.; Some tables may be marginally legible due to small size type of original document.
EDRS PRICE	MF01/PC09 Plus Postage.
DESCRIPTORS	*Bilingual Education; Elementary Secondary Education; *English (Second Language); Equal Education; *Program Effectiveness; *Program Evaluation; Second Language Instruction; *Transitional Programs
IDENTIFIERS	*Bilingual Education Act 1968; Limited English Speaking

ABSTRACT

Based on a review of twenty-eight studies, this report examines the success of transitional bilingual education programs in leading to better performance in English and in nonlanguage subject areas. The following conclusions are offered: (1) Schools can improve the achievement level of language minority children through proper programs. (2) There is not sufficient evidence for the effectiveness of transitional bilingual education to justify the Federal government's exclusive reliance on this method of instruction. Therefore, each school district should decide what type of special program is most appropriate locally. (3) Evidence does not support the necessity of teaching nonlanguage subjects in the child's native tongue, though it is necessary to structure the curriculum differently from that of English monolingual students if the subject matter is to be taught to non-English speakers. (4) Immersion programs, which involve structured curricula in English, show promising results and should be given more attention in program development. (5) The Title VII program for bilingual education must take steps to improve the quality of its program evaluations. (Author/GC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED215010

EFFECTIVENESS OF BILINGUAL EDUCATION: A REVIEW OF THE LITERATURE

Keith A. Baker

Adriana A. de Kanter

Final Draft Report
September 25, 1981

This paper was written by staff members in the Office of Planning, Budget and Evaluation. The report does not represent the official position of the U.S. Department of Education.

The authors appreciate the patience and fortitude of Sandra K. Richardson who typed and retyped many versions of this paper. Special mention should be made of William Fischer and Dr. Thomas Rhue who respectively encouraged the authors in their research. Warmest thanks must be expressed also to Dr. Alan Ginsburg and Dr. Beatrice Birman for their criticisms and assistance in each successive draft. Finally, Marshall S. Smith provided extensive suggestions about the methodological features of this paper. While the framework and perspective presented are solely the responsibility of the authors, his help has been of great value.

UD 022 052

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ABSTRACT

This report explores the effectiveness of bilingual education based on a study of the literature. The study was undertaken as part of the regulatory review of the Language-Minority Notice of Proposed Rulemaking published in August 1980 in response to Title VI of the Civil Rights Act of 1964. Although the current Administration has withdrawn the proposed rule, the need for an examination of the Department of Education's policy on the education of language-minority children continues.

This need is especially critical in light of this review's findings that the method of bilingual instruction (transitional bilingual education, or TBE) promoted by the Office of Bilingual Education and Minority Language Affairs and by the Office for Civil Rights in the Department of Education cannot be supported as the sole method used with language-minority children.

We examined well over 300 documents concerning bilingual education--which makes this the most comprehensive review to date on this subject. However, because most of the bilingual studies, especially Title VII program evaluations, are of poor quality, our conclusions have been based chiefly on only 28 studies that were methodologically applicable to our concerns.

Our conclusions, briefly summarized, are these:

- Schools can improve the achievement level of language-minority children through special programs.

- The case for the effectiveness of transitional bilingual education is so weak that exclusive reliance on this instruction method is clearly not justified. Too little is known about the problems of educating language minorities to prescribe a specific remedy at the Federal level. Therefore, while meeting civil rights guarantees, each school district should decide what type of special program is most appropriate for its own unique setting.

- There is no justification for assuming that it is necessary to teach nonlanguage subjects in the child's native tongue in order for the language-minority child to make satisfactory progress in school. However, if nonlanguage subjects are to be taught in English, the curriculum must be structured differently from the way the curriculum is structured for monolingual English-speaking students.

- Immersion programs, which involve structured curriculums in English for both language and nonlanguage subject areas, show promising results and should be given more attention in program development.

- The Title VII program for bilingual education must take steps to improve the quality of its program evaluations.

SUMMARY REPORT OF A REVIEW OF THE LITERATURE ON THE EFFECTIVENESS OF BILINGUAL EDUCATION[1]

Introduction

This investigation[2] was begun at the request of the White House Regulatory Analysis and Review Group for an assessment of the effectiveness of transitional bilingual education (TBE). The request came during that group's review of the Department of Education's proposed language-minority regulations which were issued in August 1980 in response to Title VI of the Civil Rights Act of 1964. Although the proposed rules have been withdrawn, the question of the effectivenss of transitional bilingual education is still important for the development of Department policy.

The review focused on two questions derived from the principal intent of Federal policy:

1. Does transitional bilingual education lead to better performance in English?
2. Does transitional bilingual education lead to better performance in nonlanguage subject areas?

Although a number of other goals are often recognized for bilingual education (e.g., reduced dropout rates, improved self-image and attitude toward school, preservation of the primary language and culture, and lower absenteeism), we limited this review to these two major questions. Few of the studies reviewed addressed the other goals, and a systematic assessment of their accomplishments could not be made.

This review did not directly include all the evaluations of bilingual programs that have been completed. The studies reviewed were subject to the following limitations:

- In general, we did not review studies rejected as unsound by earlier reviewers (Zappert and Cruz, 1977; Engle, 1975). An effort was made to examine all studies reported since Zappert and Cruz, the most recent prior review.
- The Office of Bilingual Education (Title VII program) was unable to provide copies of its pre-1978 evaluations, so most of them were not available. However, since Zappert and Cruz (1977) seem to have reviewed and rejected most of the pre-1978 Title VII evaluations, we believe the present review is the most comprehensive review of the effectiveness of bilingual education yet undertaken.
- Since our focus was on transitional bilingual education, our limited time and resources prohibited an equally comprehensive coverage of alternative methods. However, we have covered the major studies.

Consideration of the literature and Federal policy led to the identification of three basic instructional alternatives, in addition to the alternative of doing nothing for the language-minority child (also known as submersion):

- Submersion. Language-minority children are placed into an ordinary classroom where English is spoken. There is no special program to help them overcome the language problem. Submersion is aptly described as "sink or swim." The minority home language (L1) is not used at all in the classroom.[3]

- Structured Immersion. Instruction is in the second language (L2), as in the case of submersion, but there are important differences. The immersion teacher understands the home language (L1), and students can address the teacher in the home language (L1); the immersion teacher, however, replies only in the second language (L2). Furthermore, the curriculum is structured so that no prior knowledge of the second language (L2) is assumed when subject areas are taught. ~~All content is introduced in a way that can be understood by the students~~. The students in effect learn L2 and content simultaneously. Structured immersion differs from bilingual instruction in that the home language (L1) is never spoken by the teacher and subject area instruction is given in the second language from the beginning.

- English as a Second Language (ESL). ESL students are placed in regular instruction for most of the day. During part of the day, however, these students receive extra instruction in English. Generally, this extra help is based on a special curriculum designed to teach English as a second language. Home language (L1) may or may not be used in ESL instruction.

- Transitional Bilingual Education (TBE). Subject matter is taught in the home language (L1) until the students' second language (L2) (English) is good enough for them to participate successfully in a regular classroom. ESL is often used to help minimize the time needed to master English. Home language (L1) instruction is gradually phased out, and regular English instruction is gradually phased in. TBE is differentiated from submersion and ESL by the use of the home language (L1) for instruction.

These three instructional types sometimes shade into one another; for example, most TBE programs include an ESL component. In addition, there is a considerable range of activities incorporated within each type. Nevertheless, the typology is real and important. If the types are thought of as representing different philosophies for addressing the needs of students with limited English proficiency, it is immediately apparent that the different philosophies lead to very different classroom practices which can be identified in actual classroom settings.

The alternative instructional models differ on both civil rights and educational dimensions. Submersion is the absence of a special program which the Supreme Court found to violate the civil rights of language-minority children in the case of Lau v. Nichols. Lau was a class-action suit against the San Francisco Public School District which alleged that the district's failure to provide special educational services to non-English-speaking Chinese students violated both the equal protection clause of the 14th Amendment and Title VI of the Civil Rights Act of 1964. The Supreme Court found the San Francisco Public School District to be in violation of Title VI. However, the Court declined to prescribe a specific program that would provide equal educational benefits, stating: "Teaching English to the students of Chinese ancestry is one choice. Giving instruction to this group in Chinese is another. There may be others." Each of the three instructional methods seeks to correct the civil rights problem described by the Court in the Lau decision by providing special help to the language-minority child.

The differences among the three methods can be further illustrated in a brief outline of the underlying arguments supporting each method as a successful solution for the problems of language-minority children:

- o Transitional Bilingual Education. While children are learning English, they should be taught subject material in their home language so that their academic progress will not be retarded by their limited knowledge of English. It is easier to first learn to read in the home language than in the second language and the reading in the home language will facilitate second-language reading. Therefore, the sequence of instruction in L1 before L2 is superior to an all-second-language program.

- o English as a Second Language. Concentrated additional instruction in English language skills will keep the students from falling behind in subject areas.

- o Structured Immersion. The solution to developing English and progressing in other subjects simultaneously is to teach all subjects in English at a level understood by the students. Although the curriculum assumes no prior knowledge of English language-minority students in effect learn English as they learn math, and learn math through English instruction that is understandable at their level of English proficiency.

Methodological Approach

In reviewing a body of research to determine the effectiveness of a particular instructional program, three fundamental questions are asked:

1. Does the study present data relevant to the issues of interest?

2. Does the design of the study permit any plausible alternative explanation for the results other than that the program worked?

3. How widely can the results of acceptable studies be generalized?

The approach used in this review is based on the application of standard methodological criteria for the adequacy of research designs which are widely accepted in the education literature. These criteria are applied to the studies being examined to see if they are of acceptable quality. In this review, the criteria for methodological soundness were applied in a way that recognized that a design weakness in one area can be compensated for by strength in another area. Previous reviews have not recognized such complexities. For example, Zappert and Cruz (1977) rejected the study by Covey (1973) for failing to control for initial differences in language ability. However, because Covey's study randomly assigned students to treatment and control groups (a true experiment), no further control of language ability was necessary.

Many factors can affect the methodological quality of research and evaluation studies. Campbell and Stanley (1963), for example, list 12 broad categories, most of which contain subcategories of problems. There is general agreement in the scientific literature on what constitutes good study design. In our full report (Baker and de Kanter, 1981) we discuss the design issues encountered in each study that determine our judgment of the study's acceptability. For this review each study was assessed to determine if it addressed the relevant questions by using a methodologically sound design. The following characteristics generally led to rejection of a study:

- The study did not address our issues.

- If students were not randomly assigned to the treatment and comparison groups[4] and nothing was done to control for possible initial differences between the groups, the studies were rejected. Any differences found between the students in the special program and the group not in the special program could have been due to preexisting differences between the two groups. Differences need not have been the result of the program but of the way the groups were selected.[5]

- If studies did not apply appropriate statistical tests to demonstrate program effects, the studies were not accepted. Presenting differences between two groups is not sufficient proof that the differences did not occur by chance. Therefore, statistical tests must be introduced into the study design to verify that effects were not a chance phenomenon.

- The study used the norm-referenced design.[6] Some studies form what amounts to a control group by comparing growth against the test norms. Then these studies check to see if students in the special program showed a gain against the norm. In this design, the rate of progress of the bilingual child is compared with the rate of progress of the monolingual norming groups. There is no reason to believe that the rate of progress of bilingual and monolingual students is the same.

Therefore, any differences found by using a norm-referenced design cannot be attributed to the effects of the program.

- The study examined gains over the school year without a control group.[7] Most students learn something over the school year, so their scores will increase. If we want to know whether students gained more by being in the special language program than they would have gained in a regular school program, a comparison group of regular school students must also be included.

- The study used grade-equivalent scores.[8] Grade-equivalents do not correspond to the time pattern of learning, and the methods used to produce them are inaccurate. Equal grade-equivalent gains for two students may not represent equal learning. Use of grade-equivalents has often been criticized by evaluation experts. Quoting from "A Prototype Guide to Measuring Achievement Level and Program Impact on Achievement in Bilingual Projects" by Horst et al. (1980):

  > Grade-equivalent scores provide an illusion of simplicity but, in fact, they are almost impossible to interpret, even for specialists in test construction. Grade-equivalent scores should never be used by anyone for any purpose whatsoever.

  However, there is disagreement among testing experts whether grade-equivalents are totally unacceptable for measuring student achievement. One study that would have been accepted except for its use of grade-equivalents has been separately identified (Olesini, 1971).

By the same token, acceptable studies were—

- True experiments in which students were randomly assigned to treatment and control groups, or

- Studies using nonrandom assignment which controlled for possible preexisting differences between the groups either by matching students in the treatment and comparison groups or through using statistical procedures.

Analysis of covariance was by far the most common statistical method used to control for preexisting differences which could influence achievement between groups. Many people have serious reservations about whether this method succeeds in properly adjusting preexisting differences. Similarly, there are doubts that matching is entirely successful. For this analysis, we generally considered both methods to be acceptable unless there were defects in the application of these methods.

There are two ways results can acquire generalizability. First, the students studied can be selected in such a way as to be representative of the entire population of students in whom we are interested--in this case, language-minority children in the United States. Only 1 of the more than 300 documents we reviewed comes close to acquiring national generalizability (Danoff et al., 1977, 1978) and this study has problems. Second, generalizability comes from consistent findings in many different settings. Thus, if every study came up with the same result, no matter how limited the generalizability of each individual study, the weight of the collective evidence could be compelling. Since only one of the studies we reviewed was nationally representative, we attach great importance to finding consistency in the results of the studies when arriving at conclusions as to how well bilingual education works.

Results

Of the several hundred studies covered by the review, only 28* were found to apply to our concerns and to meet our methodological criteria. Before discussing the studies we found to be methodologically acceptable, we should note that we found several studies that have previously been widely cited as evidence for the effectiveness of TBE to be methodologically unacceptable (Skutnabb-Kangas and Toukomaa, 1976; Skutnabb-Kangas, 1979; St. John Valley, 1980; Veilleux, 1977; Leyba, 1978; Trevino, 1968; Modiano, 1968; Egan and Goldsmith, 1981; Rosier and Holm, 1980; and AIR, 1975a, 1975c, 1975e).[9]

Table 1 summarizes the 28 studies we found to apply to our two questions and to meet minimal methodological criteria; by comparison, Zappert and Cruz (1977) found 18 methodologically acceptable studies. For each study, table 1 gives the author, the grades of school encompassed, the number of students in the treatment and control groups combined, the languages used by the program, and the results the author(s) reported for second-language and math skills. The most frequent home language was Spanish, but a number of other languages were represented as well. The most common second language was English. In three studies, French was the second language. Most of the studies were neither longitudinal nor true experiments. Several studies included very large numbers of students.

For each study we examined, table 1 indicates whether the study was better than or equivalent to another approach. These comparisons were based on findings which were statistically significant. Some studies had mixed results, based either on tests or grade levels. Where mixed results are found, we have indicated the nature of the different results.

Structured immersion programs seem to have done particularly well.[10] Lambert and Tucker (1972) and Barik and Swain (1975) found second-language learning through structured immersion superior to ESL, and Pena-Hughes and Solis (1980) showed structured immersion superior to transitional bilingual education. As for nonlanguage subjects, Lambert and Tucker (1972), Barik et al. (1977), and Ramos et al. (1967) all showed that it is possible to teach math successfully in the second language. This finding suggests that if the curriculum is properly structured so that the means of communication is at a level the child can understand, there will be no negative consequences from teaching math in the second language. We found no data in

---

* Including the study by Olesini, which used grade-equivalents.

TABLE 1

SUMMARY OF APPLICABLE STUDIES

Author	Date	Grade	Design[1]	Number[2] of Students	Languages[3]		Reported Results	
					L1	L2	L2	Math
AIR (Corpus Christi)	1975b	K-1	Longitudinal; analysis of covariance	393	Spanish	English	TBE no different from submersion in 1 grade; TBE better than submersion in 1 grade	
Ames and Bicks	1978	1-9	Analysis of covariance	669	Spanish and French	English	TBE no different from ESL alone	TBE better than ESL alone
Balasubramonian et al.	1973	K-3	Analysis of covariance and other adjustments	317	Spanish	English	TBE no different from ESL alone	
Barik and Swain	1975	K-2	Longitudinal; analysis of covariance	2,253	English	French	Immersion better than ESL	Math taught in L2 no different from math taught in L1
Barik et al.	1977	2-5	Longitudinal; analysis of covariance	*	English	French		Depending on year and grade, math taught in L2 was worse than, no different from, or even better than math taught in L1
Carsrud and Curtis	1980	4-5	Longitudinal; analysis of covariance	172	Spanish	English	TBE better than submersion in 1 grade; TBE no different from submersion in 1 grade	TBE no different from submersion

(Continued)

TABLE 1. SUMMARY OF APPLICABLE STUDIES (Continued)

				Number[2] of	Languages[3]		Reported Results	
Author	Date	Grade	Design [1]	Students	L1	L2	L2	Math
A. Cohen	1975	K-3	Longitudinal analysis of covariance and other adjustments	90	Spanish	English	TBE no different from submersion on 86 of 100 language skills; submersion better than TBE on 11; TBE better than submersion on 3	TBE no different from submersion in 2 of 3 grades; TBE better than submersion in 1 grade.
Cottrell	1971	K-1	Analysis of covariance	470	Navajo	English	TBE no different from submersion	
Covey	1973	9	Random assignment	200	Spanish	English	TBE better than submersion	TBE no different from submersion
Danoff et al.	1977, 1978	2-6	Analysis of covariance and other adjustments; big study	8,900	Several	English	Submersion better than TBE	TBE no different from submersion
Huzar	1973	2-3	Random assignment; one-way analysis of covariance	160	Spanish	English	TBE no different from submersion	
Kaufman	1968	Junior High	Experiment; longitudinal	139	Spanish	English	TBE better than submersion on 2 component scores of a standardized achievement test and no different on 7 component scores in one school; TBE no different from submersion on 9 tests in another school.	
Lambert and Tucker	1972	1-4	Longitudinal; analysis of covariance	213	English	French		Math taught in L2 no different from math taught in L1

(Continued)

TABLE 1. SUMMARY OF APPLICABLE STUDIES (Continued)

Author	Date	Grade	Design[1]	Number of Students[2]	Languages[3] L1	L2	Reported Results L2	Math
Legar-reta	1979	K	Analysis of covariance	80	Spanish	English	TBE better than submersion or TBE no different from submersion, depending on the test; TBE with ESL better than TBE without ESL component	
Lum	1971	1	Random assignment	55	Chinese	English	ESL alone better than TBE on 3 tests; ESL alone no different from TBE on 2 tests	
Matthews	1979	2,4, 6,8	Log-linear model	1,011	Many	English	TBE/ESL no different from submersion	
McConnell	1980	Pre-K -3	Longitudinal; subject as own control	1,020	Spanish	English	TBE better than submersion	
McSpadden	1979	K-1	Analysis of covariance	196	French	English	TBE no different from submersion	TBE no different from submersion
McSpadden	1980	K-2	Longitudinal; analysis of covariance	263	French	English	Submersion better than TBE in 1 of 3 grades; TBE no different from submersion in 2 grades	Submersion better than TBE in 1 of 2 graded; TBE no different from submersion in 2 grades
Moore and Parr	1978	K-2	Analysis of covariance	130	Spanish	English	Submersion better than TBE	TBE no different from submersion
Olesini**	1971	3	Matching	60	Spanish	English	TBE better than submersion in 1 of 3 components of a standardized test; TBE no different from submersion in one component	TBE better than submersion on 1 component of a standardized test; TBE no different from submersion on 1 component

(Continued)

TABLE 1. SUMMARY OF APPLICABLE STUDIES (Continued)

				Number[2] of	Languages[3]		Reported Results	
Author	Date	Grade	Design[1]	Students	L1	L2	L2	Math
Pena-Hughes, and Solis	1980	K	Random assignment	156	Spanish	English	Immersion better than TBE	
Plante	1976	1-2	Longitudinal; experiment	72	Spanish	English	TBE better than submersion in 1 grade. TBE no different from submersion in 1 grade and for both grades combined	
Ramos et al.	1967	1-6	Longitudinal; matching	***	Hiligaynon Pilipino	English	TBE no different from immersion[4],****	TBE no different from immersion[4], ****
SEDL (Stebbins)	1977	K-3	Longitudinal; analysis of covariance and other adjustments; 5 sites	1,060	Spanish	English	TBE no different from submersion[4]	TBE no different from submersion
Skoczylas	1972	1	Analysis of covariance	47	Spanish	English	TBE no different from submersion	Submersion better than TBE
Stern	1975	4-6	Analysis of covariance	213	Spanish	English	Submersion better than TBE[4]	Submersion better than TBE[4]
Zirkel	1972	1-3	Matching; analysis of covariance	278	Spanish	English	TBE better than submersion on 1 test; TBE no different from submersion on 4 tests	

(Continued)

* Treatment = 73, control not given.

** Rejected for use of grade-equivalents only.

*** Unable to obtain information at present; however the sample size was large.

****The classification of the instructional method used in this study cannot be determined, but our best guess is immersion.

[1] In the case of multiyear studies, the number of tested students was counted. Rather than counting the number of unique students, the study counted each year a student was tested as a separate instance.

[2] For studies not using random assignment, we note the method used to adjust for possible preexisting differences between the treatment and control groups. Analysis of covariance is a statistical method used to adjust for preexisting differences.

[3] L1 is the language-minority child's home language; L2 is the child's second language.

[4] This result represents our conclusion from the author's very complex analysis; see chapter 2 of the full report.

these studies pertinent to other subject areas, which are often more dependent on verbal skills than math is. Ramos et al. (1967) reported the least favorable results for immersion in the literature. They found that immersion from grade 1 was as effective after 5 or 6 years as a TBE program in which all instruction was in L1 for grades 2 through 4, and in L2 thereafter.

The data on ESL instruction are not very informative. As just noted, two studies found structured immersion superior to ESL. Ames and Bicks (1978) and Balasubramonian et al. (1973) found that TBE programs which included an ESL component were no more effective than ESL alone. Lum (1971) had mixed results finding both that TBE programs which included an ESL component were no more effective than ESL alone and that ESL alone was superior to TBE. Legarreta (1979) found that a TBE program with ESL worked better than a TBE program without an ESL component.

Mixed findings were found for several of the studies. As a result, the reader will notice that there are more findings than there are studies. Mixed findings can be attributed to different achievement results either from grade to grade or between tests. Therefore, some studies may be counted more than once as showing a positive, no different, or negative finding.

With respect to TBE, positive outcomes pertaining to language performance were reported by Covey (1973), Carsrud and Curtis (1980), McConnell (1980), Olesini (1971), Plante (1976), Legarreta (1979), AIR (1975b), Cohen (1975), Kaufman (1968), and Zirkel (1972). However, the case for the effectiveness of TBE is called into question by studies that found no difference in second-language performance between treatment and comparison groups (Ramos et al., 1967; Ames and Bicks, 1978; Plante, 1976; Kaufman, 1968; Huzar, 1973; Legarreta, 1979; A. Cohen, 1975; SEDL, 1977; Carsrud and Curtis, 1980; Matthews, 1979; Skoczylas, 1972; McSpadden, 1979, 1980; Balasubramonian et al., 1973; Cottrell, 1971; Olesini, 1971; AIR, 1975b; Zirkel, 1972; Lum, 1971). Moreover, some studies found TBE to be less effective than either immersion or ESL (Lum, 1971; Pena-Hughes and Solis, 1980) and some found TBE to have negative effects by comparison with submersion (Danoff et al., 1977, 1978; Stern, 1975; Moore and Parr, 1978; A. Cohen, 1975; McSpadden, 1980).

Olesini (1971), A. Cohen (1975), and Ames and Bicks (1978) found that TBE improved acquisition of math skills. However, no effect was found by Danoff et al. (1978), Carsrud and Curtis (1980), Moore and Parr (1978), McSpadden (1979, 1980), A. Cohen (1975), Covey (1973), Olesini (1971), SEDL (1977), and Ramos (1967). Skoczylas (1972), McSpadden (1980), and Stern (1975) reported a negative effect.

Caution must be exercised in generalizing from table 1 because some issues of methodological adequacy remain. For example, Covey (1973) and McConnell (1980a, 1980b) report success for programs including TBE. However, the programs also included very low staff-student ratios--1 to 8 in the program studied by Covey (1980). Therefore, strong doubts exist as to whether the reported program effect was due to the use of bilingual instruction or to the small classes.

We also examined our findings to determine which studies would have been included if we loosened our criteria and accepted grade-equivalents. Only Olesini would then be included in our results. His results were generally favorable to TBE and have bean included in table 1 and table 2.

It is instructive to look for patterns in the findings of all these studies. Table 2 summarizes our findings with respect to comparing alternative instructional approaches. We have grouped the 28 studies according to the comparisons they examine. Then, we have aggregated their findings according to whether the study had positive, no difference, or negative results in comparison to the other approach.* For example, the first comparison in table 2 looks at the effectiveness of TBE versus submersion. For second-language acquisition, 10 findings favored TBE, 15 findings found no differences between TBE and submersion, and 5 findings were actually negative for TBE.

The results in table 2 must be qualified. Rather than simply counting the number of studies with various outcomes, we must go beyond these tabulations and give more or less weight to different findings. For example, the study by Ames and Bicks (1978) (which found that TBE produced better math results than submersion did) took place in only one school district, while the Danoff et al. (1978) study (which found that TBE had no effect on math) was designed to be nationally representative. Therefore, Danoff's findings must be given considerably more weight. Nevertheless, a clear understanding of our finding can only be obtained by looking at the studies in the aggregate rather than looking at the studies in isolation. Our policy implications are presented below.

Implications

We believe the literature makes a compelling case that special programs in schools can improve the achievement of language-minority children. There is no evidence, however, that a specific program should be either legislated or preferred by the Federal Government. Indeed, more research and demonstration projects with sound evaluation models are needed to determine which programs are effective with which types of children in which locations. The rest of this summary will present our findings.

Special Programs Can Improve Achievement in Language-Minority Students

The literature we reviewed indicates that special programs designed to overcome language difficulties in school can improve the achievement of

---

* Because, as already noted, some studies had mixed results, the reader will notice that there are more findings than there are studies. However, if a study administered five tests of which three had positive results and two negative ones, we would record only one positive and one negative result in our comparison tables.

TABLE 2

SUMMARY OF FINDINGS FROM APPLICABLE STUDIES*

Transitional Bilingual Education Versus Submersion

TBE:	Second Language	Math
Positive	10	2
No Difference	15	9
Negative	5	3

Transitional Bilingual Education Versus English as a Second Language

TBE:	Second Language	Math
Positive	1	1
No Difference	3	NA
Negative	1	NA

Transitional Bilingual Education Versus Immersion*

TBE:	Second Language	Math
Positive	0	0
No Difference	1	1
Negative	1	0

Immersion Versus English as a Second Language*

IMMERSION:	Second Language	Math
Positive	1	NA

---

* Math scores found in immersion projects in Canada are difficult to compare with scores in regular English curriculums. What can be concluded, however, is that students can achieve equally well (or better) in math classes taught in L2 as in math classes taught in L1.

language-minority children. The studies by Pena-Hughes and Solis (1980, 1981), Plante (1976), Huzar (1973), Covey (1973), Kaufman (1968), and Lum (1971) were true experiments, and all showed special programs to have positive or neutral effects. The ingenious nonexperimental design used by McConnell (1980a, 1980b) also seems to have firmly established the presence of a positive program effect. Positive effects also were reported in the nonexperimental studies of Zirkel (1972), Ames and Bicks (1968), AIR (1975b), Barik and Swain (1975), Olesini (1971), Barik et al. (1979), Lambert and Tucker (1972), Legarreta (1979), Carsrud and Curtis (1980), Cohen (1975), and Malherbe (1946). Note, though, that while special programs have been shown to be effective, this conclusion says nothing about the effects of any particular instructional approach.

The Federal Government Should Not Place Exclusive Reliance on Transitional Bilingual Education

For more than a decade, the Federal Government has worked toward institutionalizing transitional bilingual education as virtually the only approved method of instruction for language-minority children. TBE has been emphasized in Title VII funding decisions. TBE has been implemented nationwide by the Office for Civil Rights' interpretation of the Lau decision. And in 1980, the Department of Education proposed, with few exceptions, the legal mandate of transitional bilingual education through Federal regulations (a proposal that has been withdrawn by the current Administration).

When we reviewed the literature on the effectiveness of transitional bilingual education we did not find justification for such heavy reliance on this method of instruction. In order for the Federal Government to rely exclusively on one instructional method for meeting the needs of language-minority children, the following two conditions must hold:

1. There must be a strong case that the instructional method is uniformly effective.

2. Effective instructional alternatives should not exist. If the desired outcomes can be reached through more than one approach, the Federal Government should not constrain the options of local schools.

Only 28 studies that passed our methodological test addressed the effectiveness of TBE, and only 11 of the 25 studies looking at TBE reported a positive effect. Further, additional methodological problems in these studies impose strong limits on generalizing their results. Three studies suggest that the reported positive outcome could well have been due to other aspects of the program rather than to TBE itself (Covey, 1973; McConnell, 1980a, 1980b; Plante, 1976). In addition, a number of studies that used multiple-outcome measures found mixed results. Several other studies found a negative effect for TBE when compared with submersion, ESL, or immersion (Danoff et al., 1977; Moore and Parr, 1978; McSpadden, 1980; Skoczylas, 1972; Cohen, 1975; Lum, 1971; Stern, 1975; Pena-Hughes and Solis, 1980). Although we reviewed a limited number of immersion studies, each analysis of structured immersion generally found positive findings for that approach. Achievement in both language skill and subject matter knowledge was better through structured immersion than through ESL or TBE (Barik and Swain, 1975; Barik et al., 1977; Lambert and Tucker, 1972; Pena-Hughes and Solis, 1980).

These findings do not add up to a very impressive case for the effectiveness of transitional bilingual education. We conclude that TBE fails both tests for justifying reliance on it as the exclusive method for instructing language-minority children. There is no firm empirical evidence that TBE is uniquely effective in raising language-minority students' performance in English or in nonlanguage subject areas.

Since several States have followed the Federal lead in developing programs for language-minority children—in some cases, even legislating TBE—our analysis has implications beyond the Federal level.

Federal Policy Should Be Flexible

For more than a decade, Federal policy (as expressed through Title VII legislation, Title VII funding decisions, OCR implementation of the "*Lau* Remedies," and the August 5 Notice of Proposed Rulemaking) has emphasized transitional bilingual education to the virtual exclusion of alternative methods of instruction. We found through our analysis that this policy is not justified on the basis of educational effectiveness. While transitional bilingual education has been found to work in some settings, it has also been found ineffective and even harmful in other places. Furthermore, both of the major alternatives to TBE—structured immersion and ESL—have been found to work in some settings.

The commonsense observation that children should be taught in a language they understand does not necessarily lead to the conclusion they should be taught in their home language. They can be successfully taught in a second language if it is done right. The key to successful teaching in the second language seems to be to insure that the second language and subject matter are taught simultaneously so that subject content never gets ahead of language. Given the American setting, where the language-minority child must ultimately function in an English-speaking society, carefully conducted second-language instruction in all subjects may well be preferable to bilingual methods.

We conclude that it is very hard to say what kind of program will succeed in a particular school. Hence it seems that the only appropriate Federal policy is to allow schools to develop instructional programs that suit the unique needs and circumstances of their students.

There is no reason to assume a priori that the same approach that is applied to a rural Southwest Texas district with a large proportion of second-generation Hispanic children should also be applied to a district with a small group of Lao refugees in a Northern city. But Federal policy has been based on such an assumption over the years. Our review indicates that a fundamental change in Federal policy is needed.

We believe this change will require recognition by the Department of Education that other pedagogical methods for language-minority children can be effective and can meet civil rights criteria. Federal funding practices must encompass each of the special programs designed to meet the needs of language-minority children so that a more realistic balance among various program types is achieved.

A widespread structured immersion demonstration program is especially needed. Until now, the immersion method has been rejected on the basis of weak theoretical arguments.[11] Immersion may not transfer successfully from Canada to the United States, but this is an empirical question that should be answered by direct test. As a first step, the Department should immediately fund an extensive evaluation of the McAllen, Texas, program, which has a true experimental design for comparing the effectiveness of structured immersion and TBE for Mexican-American students of low socioeconomic status.

Given the complexity of the problem, it also seems that the Federal Government should provide the most current information on pedagogical methods for language-minority children so that school districts can make informed choices, adapting methods to their local needs.

Improved Bilingual Research and Program Evaluations Are Needed

More and better research and improved program evaluations in bilingual education are necessary if the needs of language-minority children are to be adequately met. The low quality of the methodology found throughout the literature is a serious problem. The major methodological problems with the literature include the following:

- The absence of random assignment between treatment and control groups,
- The use of study designs that cannot show a treatment effect in the absence of random assignment, such as the norm-referenced model or failure to use analysis of covariance, and
- The failure to apply appropriate statistical tests to demonstrate program effects.

These problems have particularly characterized Title VII evaluations. The Title VII bilingual program has begun to take steps to improve the quality of local results. However, our review has indicated that program evaluations are still of very poor quality; much improvement is still needed in this area.

Bilingual education involves many complex, difficult issues that have been little (or insufficiently) studied. Federal funding for research in the area of bilingual education was allotted for the first time under Part C of Title VII in 1978, with the Elementary and Secondary Education Act amendments (ESEA). The need for additional research is great.

Unfortunately, however, when Congress established the legislation in 1978, it limited research to examining transitional bilingual education specifically, rather than all pedagogical methods for students with limited English proficiency. As a result, Federal research has been skewed to focus on one method. Ultimately, the development of effective instructional programs for language-minority children will come about only through a more broadly based research agenda.

Areas for redirected research should include the following:

- A study of the divergent educational needs of language-minority children in the United States to include the examination of how these children's language deficiencies differ in their home language and English,
- Examination of the effectiveness of alternative instructional approaches and how these approaches meet the needs of different types of language-minority children,
- A reexamination of the theory of TBE (designed for monolingual L1 speakers), which may not be relevant to many of the language-minority students in the United States,
- Formulation of appropriate structured immersion curriculums,
- Examination of the methods of English as a second language (vocabulary drills versus meaningful English communication), and
- Examination of bilingual education teacher qualifications and the degree of fluency such teachers have in both languages.

NOTES

1. The full report is found in Baker and de Kanter (1981).

2. A literature review is a secondary analysis limited by the level of detail the authors provide. This limitation was taken into account when drawing conclusions from the literature.

3. L1 refers to the child's first, or home, language. L2 refers to the second language, the language used by society and in the schools (in the case of language-minority students). In the United States, L2 refers to English for the language-minority child; L1 is that child's home language (Spanish, Chinese, etc.).

4. Examples of studies rejected because they used nonrandom assignment, posttest-only design are these: South San Francisco (1979), Elligett (1980), AIR (1975), JDRP (1977b).

5. Among the factors affecting the performance of language-minority children in school, especially in learning English, are *age* (Krashen, 1979; Asher and Garcia, 1969; Giles, 1971), *differences in learning between oral and written language skills* (Cummins, 1978, 1980; Fishman, 1965), *socioeconomic status* (Moore, 1978; Veltman, 1980; Rosenthal et al., 1978, de Avila, 1981), *various cognitive abilities* (Darcy, 1953; Peal and Lambert, 1962; Landry, 1974; Segalowitz, 1975; Humphrey, 1977; Coronado, 1979; Malherbe, 1946; Fishman, 1965; Jensen, 1963a, 1963b; Johnson, 1953, cited in Albert and Obler, 1980), *place of birth*—immigrant or native-born (Carter, 1970; Troike, 1978; Kimball, 1968; Anderson and Johnson, 1971; Cardenas and Cardenas, 1972; Baral, 1979; Ferris, 1979), *the degree of home language dominance* (Battel et al., 1975), ethnicity (Rosenthal et al., 1981; Matthews, 1979; Veltman, 1980; Balasubramonian et al., 1973; Baker and de Kanter, 1981), *the student's motivation and self-concept* (Christian, 1976; Modiano, 1973; Zirkel, 1972, von Maltitz, 1975; Del Buono, 1971; Skoczylas, 1972; Rand, 1980), *parental support for the educational program* (Lambert and Tucker, 1972; Del Buono, 1971), and *characteristics of the community* (Lambert and Sidoti, 1980; Lambert and Tucker, 1972; Skoczylas, 1972; Read, 1980).

In addition to these background characteristics of the child, numerous factors associated with the school and the educational program can affect the outcome of bilingual instruction (McDonald and Elias, 1976; Dulay and Burt, 1979; Engle, 1975; Patniz et al., 1976; de Kanter, 1979; Kramer, 1979).

6. The following studies are examples of uses of a norm-referenced design: Rimm (1980), Young (1980), Stern (1975), Corpus Christi (1980a, 1980b), JDRP (1977b), St. John Valley (1980), AIR (1975), Ames and Bicks (1978), Arce (1979), Fairfax County (1980).

7. The following are examples of studies that examined gains without any control groups: Guerrero (1980), Liberty Union (1980), Birmingham (n.d.), Valencia (1971), Smith (1978), Cahill and Foley (1973), Ghini (1979), B. Cohen (1971), Stern (1975), Price (1978), South San Francisco (1979), Charlotte-Mecklenberg (1980), Arce and Sosa (1975).

8. Examples of the use of grade-equivalent scores can be found in the following studies: Educational Management Services (1976), Young (1980), External Evaluators (1979), Giles (1971), Trevino (1970), Kaufman (1968), Olesini (1971), Leyba (1978), Hanson (1980), Elligett (1980), South San Francisco (1979), Charlotte-Mecklenberg (1980), JDRP (1977b), St. John Valley (1980), AIR (1975), Arce (1979).

9. See Baker and de Kanter (1981), chapter 3, for a detailed discussion of the methodological problems found in these studies.

10. Proponents of TBE have raised questions about the generalizability of the immersin studies based on middle-class Canadian children (see Tucker, 1980). Genesee (1976) reviewed the status of the Canadian literature and concluded that immersion was applicable to children of lower socioeconomic status and to minorities. Pena-Hughes and Solis (1980) certainly indicates that immersion is workable in the United States, but more research needs to be done because this question ultimately is an empirical one.

11. See Tucker (1980) for a complete discussion of the inapplicability of immersion to the situation in the United States and Genesee (1976) for a discussion of the generalizability of immersion to the Canadian situation. Pena-Hughes and Solis (1980) discuss a successful structured immersion program in the United States with language-minority children of low socioeconomic status.

CHAPTER 1

CONSIDERATIONS IN REVIEWING THE LITERATURE

This investigation was begun at the request of the White House Regulatory Analysis and Review Group for an assessment of the effectiveness of transitional bilingual education (TBE). The request came during that group's review of the Department of Education's proposed bilingual regulations issued in August 1980. Although the new Administration withdrew the proposed rules, the question of the effectiveness of transitional bilingual education for language-minority students is still relevant for the development of Department policy. It is also a major issue for several States whose bilingual programs have followed Federal precedent.

Since a comprehensive review of the literature on the effectiveness of transitional bilingual education has not been done for 5 to 6 years (Zappert and Cruz, 1977), a current review was necessary to meet the policy needs of the Federal Government. Such a review is a complex task, requiring not only the identification of studies bearing on our questions but also the application of standards of scientific research. These standards provide a measure of the methodological adequacy of each study's approach and the extent to which the study results can be generalized into Federal policy recommendations. To enable the reader to follow how we carried out this task and to judge the validity of our conclusions, we have carefully defined our methodological criteria and illustrated how we applied these criteria in the review procedures.

In the remainder of this chapter, we will summarize the history of the Federal Government's involvement in the issue of the civil rights of language-minority children, describe alternative methods of instruction, explain how the review was done, and discuss the major methodological problems found in the studies.

History of Federal Involvement in Civil Rights Issues for Language-Minority Students

Data collected by the Federal Government and private civil rights and educational organizations in the late 1960's revealed substantial evidence of discrimination against language-minority students, especially Hispanics, in the Nation's public elementary and secondary schools. Statistics on academic achievement and school retention documented the effects of this discrimination. It was clear that hundreds of thousands of language-minority students suffered severe academic retardation and exceptionally high dropout rates.

While conducting compliance reviews, the Office for Civil Rights of the Department of Health, Education, and Welfare (DHEW) discovered a number of common practices that had the effect of denying equal educational opportunities to language-minority students. These practices related to the way in which schools responded to the English language skill deficiencies characteristic of many language-minority students.

On the basis of this evidence, and using its authority under Title VI of the Civil Rights Act, the Office for Civil Rights sent a memorandum to school superintendents on May 25, 1970, "to clarify DHEW policy on issues concerning the responsibility of school districts to provide equal educational opportunity to national origin/minority group children deficient in English language skills." The memorandum stated in part:

> Where inability to speak and understand the English language excludes national origin/minority group children from effective participation in the educational program offered by a school district, the district must take affirmative steps to rectify the language deficiency in order to open its instructional program to these students.

The memorandum required that--

- School districts not use English language ability as a basis for assigning national origin/minority group students to classes for the mentally retarded or to deny these students access to college preparatory courses "on a basis directly related to the failure of the school system to inculcate English language skills."
- "Any ability grouping or tracking system employed by the school system to deal with the special language skill needs of national origin/minority group children must be designed to meet such language skill needs as soon as possible and must not operate as an educational dead-end or permanent track."
- School districts must notify the parents of national origin/minority group students of the school activities that are called to the attention of other parents. The notice, to be adequate, must be in a language they can understand.

Lau v. Nichols Case

The Office for Civil Rights memorandum was affirmed by the Supreme Court in its 1974 decision in the case of Lau v. Nichols. Lau was a class-action suit against the San Francisco Public School District which alleged that the district's failure to provide special educational services to non-English-speaking Chinese students violated both the equal protection clause of the 14th Amendment and Title VI of the Civil Rights Act of 1964.

In Lau, the Court reviewed the California Education Code and concluded that--

> Under these State-imposed standards there is no equality of treatment merely by providing students with the same facilities, textbooks, teachers, and curriculum; for students who do not understand English are effectively foreclosed from any meaningful education.

> Basic English skills are at the very core of what these public schools teach. Imposition of a requirement that before a child can effectively participate in the educational program, he must already have acquired those basic skills is to make a mockery of public education. We know that those who do not understand English are certain to find classroom experiences wholly incomprehensible and in no way meaningful.
>
> It seems obvious that the Chinese-speaking minority receives less benefits than the English-speaking majority from respondents' school system which denies them a meaningful opportunity to participate in the educational program—all earmarks of the discrimination banned by the Regulations.

The Court declined to rule on the constitutionality of the school district's program, focusing instead on the statutory prohibition against national origin discrimination set out in Title VI of the Civil Rights Act of 1964. Accordingly, the Supreme Court found the San Francisco Public School District to be in violation of Title VI. However, the Court declined to prescribe a specific program that would provide equal educational benefits, stating:

> Teaching English to the students of Chinese ancestry is one choice. Giving instruction to this group in Chinese is another. There may be others.

The precedent of relying only on transitional bilingual education to meet the legal requirement was established by the Federal Government in implementing Title VI of the Civil Rights Act.

Lau Remedies

Following the Supreme Court's decision in Lau, the Department of Health, Education, and Welfare assembled a group of outside and departmental education experts to develop informal policy guidelines outlining the remedial responsibilities of school districts that failed to comply with Title VI and the principles enunciatd in Lau. This group produced a document entitled "Task Force Findings Specifying Remedies Available for Eliminating Past Educational Practices Ruled Unlawful Under Lau v. Nichols"—better known as the Lau Remedies.

The Lau Remedies outlined the major elements that should be included in school districts' plans to remedy Title VI Lau violations. According to the Lau Remedies, compliance plans should, among other things, provide for the following:

- Identification of students with a primary or home language other than English.

- Assessment of the relative proficiency of such students in English and their native language.

- Instruction of elementary students through their strongest language until the students are able to participate effectively in a classroom where instruction is given exclusively through English.

- Provision of special language instruction and compensatory educational services to secondary school, language-minority students who are underachieving academically.

Because the Lau Remedies were never published as proposed regulations, their underlying assumption that TBE was the best, if not the only, way to satisfy the civil rights requirements was never opened to public debate.

A cover letter transmitting the Lau Remedies to school officials explained the document's legal application. In part it stated:

> Voluntary compliance plans which set forth educational strategies consistent with the approaches outlined in this document and which contain the other elements specified therein, will be accepted by this office. School districts submitting voluntary compliance plans to this office which are not consistent with the outlined approaches or with other required plan elements must demonstrate affirmatively, at time of submission, that such plans, at a minimum, will be equally effective in ensuring equal educational opportunity.

Thus, the Federal Government placed the burden of proof on the schools to demonstrate that an alternative to TBE was effective, even though the Government had never shown TBE to be effective.

Although DHEW used the Lau Remedies to negotiate numerous voluntary compliance plans, the document's legal authority was challenged in a 1978 Federal court suit, Northwest Artic v. Califano. As a result of the suit, DHEW agreed to publish its Title VI Lau compliance standards in the Federal Register for public comment. In keeping with the court-approved agreement, the Department of Education published a Notice of Proposed Rulemaking August 5, 1980.

The standard proposed in the notice required that transitional bilingual education be used to meet the needs of all language-minority students identified as eligible for services. It called for special instruction to encourage fluency in English while other content areas of the curriculum would be taught in the child's home language until that child mastered English well enough to succeed with all instruction in English.

Transitional bilingual education is only one of several instructional methods under the generic rubric of bilingual education that attempt to meet the needs of language-minority students. Because alternative instructional approaches are available for meeting the educational and civil rights needs of these children, transitional bilingual education should clearly have been known to be superior to the alternatives when the Federal Government attempted to mandate use of this method in the schools.

Following an extended period of public comment on the proposed regulations, the notice was formally withdrawn February 2, 1981. Although the proposed rules were withdrawn, the question of the effectiveness of transitional bilingual education is still important for the following reasons:

- The Department may still be under a consent decree requiring that the Lau Remedies be replaced with formal regulations, so new regulations may have to be drafted. If a particular instructional approach can be justified, mandating it would be one option considered in the development of new regulations.

- Withdrawal of the Title VI Language-Minority Notice of Proposed Rulemaking and the absence of an alternative leaves the Lau Remedies in force. The past practice of the DHEW Office for Civil Rights (now in the Department of Education) in developing compliance agreements has stressed transitional bilingual education through the Lau Remedies. Therefore the issue of the effectiveness of transitional bilingual education is as appropriate in assessing the Lau Remedies as in assessing the Notice of Proposed Rulemaking.

- Other Departmental policies and programs—especially Title VII funding practices—have strongly emphasized transitional bilingual education to the exclusion of alternative methods of instruction. In fact, the Department's whole approach to the problem of the language-minority child over the past decade can fairly be characterized as being based on the assumption that TBE is, with rare exceptions, the only acceptable approach. It is worthwhile to reassess this assumption by reviewing the evidence on TBE effectiveness.

- Federal policy has formed the basis for bilingual programs and legislation in several States. When States follow the Federal lead, both the States and the Federal Government need to be sure that the path taken by Federal policy is justified.

Types of Programs for Language-Minority Students

A number of schemes for classifying types of bilingual instruction exist; Valencia (1969), for example, identified 19 different models of bilingual education (also see Paulston, 1975). For our purposes three models identified in the literature are sufficient to compare with the "submersion" method—that is, doing nothing for the language-minority child.

In a submersion program, language-minority children are placed in an ordinary, English-speaking classroom with no special program to help them overcome the language problem. Submersion is aptly described as "sink or swim." The child's home language (L1) is not used in submersion.* The Supreme Court's Lau decision in effect outlawed submersion programs in the United States.

The three alternative methods that are not ruled out by the Lau decision are the following:

1. English as a Second Language (ESL). In an English as a second language program, language-minority students are placed in regular instruction for most of the day. During some part of the day, however, their curriculum differs from that of the regular classroom in giving extra instruction in mastering English. Generally, this extra help is based on a special curriculum designed to teach English as a second rather than as a first language. L1 may or may not be used in ESL instruction.

2. Transitional Bilingual Education (TBE). Subject matter in a transitional bilingual education program is at least partially taught in the home language (L1) of language-minority children until their second language (L2) (English) is good enough for them to participate successfully in a regular classroom. English as a second language is often used in conjunction with a transitional bilingual education program to help minimize the time the children spend in mastering English. It is also generally held that learning to read in L1 facilitates learning to read in L2. Sometimes L1 instruction is gradually phased out and regular English instruction is gradually phased in; in other cases, the change is more abrupt, with students being mainstreamed out of the L1 program. The ultimate goal of transitional bilingual education is to move students into the all-L2 program. This method is differentiated from the submersion and English as a second language methods by the use of L1 for instruction.

3. Structured Immersion. In a structured immersion program, almost all instruction is given in L2. There are, however, fundamental differences between structured immersion and submersion. First, immersion teachers are bilingual. Second, students can ask questions of the teacher in L1, although the teacher generally replies only in L2. Most important, the curriculum is structured so that it does not assume prior knowledge of the second language when subject areas are taught. All content is introduced in a way that can be understood by the students. The students, in effect, learn L2 and content simultaneously. Immersion differs from bilingual instruction in that it deemphasizes home language use by the teacher and it gives subject area instruction in L2 from the beginning. Structured immersion programs may include a period of L1 language arts during the school day.

---

* L1 is an abbreviation for the child's first learned language. L2 is the child's second language. In the case of language-minority groups in which some children could have learned both languages from birth, L1 refers to the non-English (minority culture) language and L2 is the normal language of schooling and of the majority culture (i.e., English in the United States).

These three program descriptions are, of course, ideal types. In practice, it is sometimes hard to classify a particular program as one or another. Furthermore, each of these general models can encompass a considerable range of activities. Nevertheless, these types have real and significant distinctions.

Perhaps the best way to conceive of these models is as different educational philosophies that guide the development of specific programs. What is considered appropriate or inappropriate for a program will vary, depending on which one of these philosophies is followed. The models correspond to real program distinctions recognized in the literature, since we derived them largely from the way the literature conceptualizes programmatic differences.

Before proceeding, we should note that the three instructional models and submersion differ on both civil rights and educational dimensions. Submersion is the lack of a special program which the Supreme Court found to violate the civil rights of language-minority children. The other three methods are alike in seeking to correct the civil rights problem described by the Court in the _Lau_ decision by providing special help to language-minority children.

The differences among the three methods can be illustrated by an outline of the arguments advanced as to why each method should succeed in solving the problems of language-minority children:

- _English as a Second Language_: Concentrated, supplementary instruction in English language skills will enable students to learn English fast enough to keep up with English-speaking teachers in the various subjects.

- _Transitional Bilingual Education_: While the children are learning English, they should be taught subject material in their home language so their academic progress will not be retarded by their lack of English skills. It is easier for the language-minority child to first learn to read in the home language rather than in English. Further, first learning to read in the home language will facilitate learning to read in English.

- _Structured Immersion_: The solution to developing students' English while they progress in other subjects is to teach all subjects in English at a level understandable to the students. Although the curriculum assumes no prior knowledge of English, language-minority students in effect learn English as they learn math and learn math through English instruction that is understandable at their level of English proficiency.

The Plan of This Review of the Literature[1]

Although many of the studies we reviewed examined several outcome measures, we are limiting the discussion in this report to two questions: Does bilingual education lead to better performance in English? Does it lead to better performance in nonlanguage subject areas? Our decision to concentrate on English and subject matter acquisition stems from basic Federal policy that recognizes the need to prepare language-minority children to function successfully in an English-speaking nation, and seeks to provide equal educational opportunity to them.

A program that produces mediocre English performance while maintaining the home language skills will be judged a worse program than one that produces better second language performance while ignoring home language skills. The justification for this viewpoint is that, in the United States, any successful education program must prepare the students to participate in an English-speaking society. Therefore, the overriding concern in evaluating instruction for bilingual students is how well they learn English.

A number of other goals have been put forward for bilingual education: less absenteeism, lower dropout rates, improved self-concepts and attitudes toward school, and development of bilingual adults. Any attempt to systematically address all these goals is beyond the scope of this report. We will, however, from time to time call the reader's attention to certain interesting findings regarding these other goals.

This is the most comprehensive review of the literature on bilingual education to date. Studies were identified by ERIC search, by consultation with experts in the field, from prior reviews, and from lists of studies. In each study we reviewed, we looked at the bibliography in an effort to identify additional studies.

We reviewed more than 300 documents (see attached bibliography). Of these, about 150 were program evaluations. In addition, Zappert and Cruz (1977) reviewed 175 studies, mostly Title VII evaluations which the Title VII program office can no longer produce. We did manage to locate most of the 12 methodologically acceptable studies Zappert and Cruz cited, and we accept their judgment that the rest were unsound.

Most, but not all, of the studies we reviewed are Title VII evaluations. They cover every region of the country, almost every State, rural and urban areas, migrant and nonmigrant students, and a variety of language groups. Some studies come from other countries.

Methods Used to Assess the Studies

Once we determined the focus of the review, we read some 300 documents looking for answers to our questions. A study may be of no use in answering these questions for either of two reasons: First, it may be looking at different questions, and so it simply does not apply to our concerns. A number of studies address specific local issues that do not generalize to our questions. Therefore, we were unable to make any further use of these studies. Second, flaws in a study's methods may raise doubts as to whether the reported program outcome might have been the result of something other than the intended effect of the program. Thus, the initial review task was to decide which studies are applicable to the issue and which studies cannot be used because of methodological problems.

The approach used in this review was based on the application of standard methodological criteria for the adequacy of research designs which are widely accepted in the education literature. These criteria were applied to the studies being examined to see if they were of acceptable quality. In this review, the criteria for methodological soundness were applied in a way that recognized that a design weakness in one area can be compensated for by strength in another area. Previous reviews have not recognized such complexities. For example, Zappert and Cruz (1977) rejected the study by Covey (1973) for failing to control for initial differences in language ability. However, because Covey's study was a true experiment, no further control of language ability was necessary.

The list of possible methodological pitfalls is long, and we did not require that a study pass every hurdle. Rather, we looked for fatal flaws in the study. A fatal flaw can be a single problem, such as using a posttest-only design without random assignment. The flaw also may be the cumulative impact of a number of separate problems.[2]

The basic objective of scientific research is to rule out alternative explanations. Scientific research is easier to describe in theory than it is to put into practice, however, and few, if any, studies succeed in completely overcoming all possible methodological problems and in eliminating all possible alternative explanations. The reviewer must exercise professional judgment as to whether there is a reasonable possibility that the authors of the study ruled out alternative explanations.[3]

For this review, the methodological problem was to be sure that the observed results were the consequence of the program under study and that alternative explanations for the results can be ruled out by virtue of the study design. The best method for achieving this goal is an experiment in which random assignment is used to select two groups of students, one of which receives the usual school program while the other receives the special treatment (instruction in English as a second language, immersion, or transitional bilingual education). Performance of the two groups is then compared by statistical methods which enable us to estimate the probability that the observed differences between the two groups were not due to chance.

Although experiments are rare in educational research, we did locate six true experiments in our literature search. In the absence of a true experimental design, the problem of carrying out a study that eliminates alternative explanations becomes more difficult, since there are a variety of sources of possible alternative explanations. The methodological glossary in appendix A lists many of the possible methodological problems that might be encountered.

We found the following kinds of studies to be clearly acceptable:

- True experiments (with random assignment between treatment and control groups).
- Studies using nonrandom assignment in which researchers had made some arrangements, either through matching or through statistical adjustment (generally analysis of covariance), to control for possible preexisting differences between the groups. (See Lambert and Tucker (1972) and McConnell (1980a, 1980b) for examples of how good design can overcome the problems of nonrandom selection.)

It should be noted that we did not automatically discount all nonexperimental studies. We recognize that both matching and analysis of covariance have been severely critized for failing to overcome the problems created by nonrandom selection. While these criticisms are certainly valid, at least in the sense that true experiments are clearly superior to the alternatives, ideal conditions are seldom met in the real whe real world of educational evaluation. Matching and analysis of covariance are generally accepted methods of correcting for the problem of nonrandom selection, other things being equal. Therefore, we do not insist on considering only true experiments as methodologically sound.

Our extensive methodological assessment determined the limits of generalizability of each study's results and the implications that could be drawn from them. Obviously, a study's results apply to the particular group of students studied, but this information alone is not very useful. Ultimately we want to know if the conclusions apply to all language-minority students or only to some particular segment.

Results can acquire wide generalizability in two ways. First, the students studied can be selected in such a way as to be representative of the entire population of students we are interested in--in this case, language-minority children in the United States. Only 1 of the almost 300 documents we reviewed falls into this category (Danoff et al., 1977, 1978). Second, consistent findings in many different settings can be the basis for generalizing. Thus, if every study came up with the same result, no matter how limited in generalizability each individual study was, the weight of the collective evidence would be compelling.

Reasons for Rejection of Studies

We considered any of the following characteristics sufficient cause to reject a study as unsuitable for our purposes:

1. Failure to address the issues we are considering here,
2. Nonrandom assignment with no effort to control for possible initial differences between control and program groups,
3. Norm-referenced design,[4]
4. Comparison of "posttest" scores only, with nonrandom assignment,[5]
5. Reliance on school-year gains for the program group without a control group, or [6]
6. Reliance on grade-equivalent scores.[7]

We have already discussed the first item. Discussion of the remaining items follows.

Nonrandom Assignment Without Control for Possible Initial Group Differences. The basic problem in assessing a study in which random assignment was not used is to insure that the group exposed to the treatment does not differ from the control group on some other variable that affects learning. Selection bias is a possible consequence of nonrandom assignment of pupils to the control and program groups; the supposed observed outcome of the program may simply result from original differences between the two groups on some characteristic related to achievement.

Our first step, therefore, was to identify what other factors are known to affect the learning process of bilingual students. Among the factors affecting the performance of language-minority children in school, especially in learning English, are the following:

- Age (Krashen, 1979; Asher and Garcia, 1969; Giles, 1971);
- Differences in learning between oral and written language skills (Cummins, 1978, 1980; Fishman, 1965);
- Socioeconomic status (Moore, 1978; Veltman, 1980; Rosenthal et al., 1978; de Avila, 1981);
- Ethnicity (Rosenthal et al., 1981; Matthews, 1970; Veltman, 1980; Balasubramonian et al., 1973; Baker and de Kanter, 1981);
- Student's motivation and self-concept (Christian, 1976; Modiano, 1973; Zirkel, 1972; van Maltitz, 1975; Del Buono, 1971; Skoczylas, 1972; Rand, 1980);

- Parental support for the educational program (Lambert and Tucker, 1972; Del Buono, 1971);

- Characteristics of the community (Lambert and Sidoti, 1980; Lambert and Tucker, 1972; Skoczylas, 1972; Read, 1980)[8];

- Various cognitive abilities (Darcy, 1953; Peal and Lambert, 1962; Landry, 1974; Segalowitz, 1975; Humphrey, 1977; Coronado, 1979; Malherbe, 1946; Fishman, 1965; Jensen, 1962a, 1962b; Johnson, 1953, cited in Albert and Obler, 1980);

- Place of birth—immigrant or native-born (Carter, 1970; Troike, 1978; Kimball, 1968; Anderson and Johnson, 1971; Cardenas and Cardenas, 1972; Baral, 1979; Ferris, 1979); and

- Degree of home language dominance (Battel et al., 1975).

In addition to these background characteristics of the child, numerous factors associated with the teacher, school, and the education program can affect the outcome of bilingual instruction (McDonald and Elias, 1976; Dulay and Burt, 1979; Engle, 1975; Patniz et al., 1976; de Kanter, 1979; Kramer, 1979).

The procedures used to assign students to bilingual programs can introduce bias according to student characteristics. For example, bias occurs when parents are permitted to volunteer their children (nonrandom assignment) for a special bilingual program. Parents who volunteer children are usually more involved in their children's schooling than parents who do not volunteer their children. They may provide more help and encouragement to the children in their school work than do the parents who remain silent. Moreover, superior students are likely to come from a home environment in which the parents are actively involved with their children's schooling. We must suspect that volunteered students are likely to be better students than other children are. Thus, the program may show "gains" due to the inclusion of better students even though the program is in reality no more effective than regular schooling (see Laumann, 1969).

Another possible bias introduced with volunteered students is that parents of children with an unusual gift for languages may want those children to benefit from a special language program. Again, students' progress may have little to do with the specific program—rather, gifted students would stand out in any language program.[9]

Graduation from the program also may introduce a selection bias. If schools mainstream students as they reach some level of performance, students who perform relatively poorly will accumulate in the program for two reasons:

- Poor performers will stay in the program longer.

- The better performers who leave will, on the average, be replaced with students who are poorer performers than the graduates were.

Comparison With Norms. When the effects of a program are evaluated, the performance of students in the program must be compared with the performance of a similar group of children not in the program. If random assignment is not used, it becomes very difficult to insure that the comparison group is similar. One approach to the problem is the norm-referenced model, in which the performance of students in the program is compared against national norms by measuring fall to spring gains in percentile scores. This is by far the most commonly used model in Title I evaluation (Goor et al., 1980). However, the applicability of the norm-referenced model in the evaluation of bilingual programs is questionable.

When the norm-referenced model is used in an evaluation, it is assumed that the expected rate of improvement of students in the program would have been the same as that of the norming group in the absence of the special program. For several reasons, this assumption is probably not valid for language-minority children.

The nature of the learning curve for language-minority children is not known. It is often assumed (see Egan and Goldsmith, 1981) that, in the absence of special help, these children will fall further behind the norms over time, since they cannot understand instruction as well as do the monolingual English-speaking children upon whom the norm is based. We found reason to question this assumption, however. Using a nationally representative sample, we found that when standardized vertical-scale scores were examined over a 3-year time period the language-minority students began below the monolingual English-speaking group but did not fall further behind over the 3 years. Although more research is needed on this point, it calls into question the assumption that the performance of language-minority children gets worse as they mature (compared with the performance of monolingual English-speaking students).

A second problem of the norm-referenced model is that, as the monolingual non-English-speaking limit is approached, an achievement test becomes both a test of communication in English and an achievement test. If students know the answers to the questions but cannot understand the test, their scores will be low. If they then learn enough English to be able to understand the test, their scores will rise because they now can communicate to the test what they know. Therefore, they will register large gains on the test even though they have not increased their knowledge of what the test purports to be measuring. It is possible, therefore, that small increases in English skills will translate into large gains on the test for the initially lowest scoring students (evidence of this phenomenon was found in Garcia, 1979; Young, 1980; Cohen, 1975; Stern, 1975).

We believe this phenomenon accounts for the spectacular gains of percentile scores, especially in math, occasionally reported for bilingual students. It is not so much that they learned better when instructed in the home language as it is that they learned enough English during the school year to be able to communicate to the test what they already knew. If this analysis is correct, then any use of the norm-referenced model in evaluating bilingual programs is highly suspect.

Finally, in the Lau decision the Supreme Court ruled that submersion is an unacceptable educational approach for language-minority children. There is no question that submersion students will eventually learn English. The issue is that it takes too long and that more effective methods of English instruction should be employed. Therefore, the appropriate standard of comparison in bilingual program evaluations is the performance of submersion students, not the monolingual English-speaking norm. For these reasons, we decided norm-referenced designs were not acceptable in addressing the issues with which we are concerned.

Comparison of Posttest Scores Only, With Nonrandom Assignment. Some studies compare only posttest scores of students in the program and of a nonrandomly selected comparison group. This design is acceptable in a true experiment, since the random assignment of students insures the comparability of the experimental and control groups. However, if assignment was not random, this design does nothing to take into consideration preexisting differences that could lead to differential performance between the experimental and control groups. Therefore, this approach does not constitute an acceptable evaluation design.

School-Year Gains Only, Without a Control Group. Several evaluations report only the differences between the program students' fall and spring scores, even testing the gain for significance. This procedure is unsound. Almost all students show some absolute gain over time, even if they are at the same time rapidly falling behind the norm. Consider the following hypothetical data:

	SCORE		
	Fall	Spring	Gain
Program.........	100	125	25
Control.........	100	150	50

According to evaluations that consider only program students, the gains would indicate program success. It is clear, however, that since the control group gained more than the program students, the program was far from effective. The point is that a simple examination of gains over the school year for students in a special program yields too little information to permit determining whether the program worked. Researchers must also compare the progress of program students with the normal rate of progress made by students not in the special program. Hence, a study design that examines gains over the school year without a control group is unacceptable.

Grade-Equivalent Scores. Studies that are based on grade-equivalent scores pose serious problems. Grade-equivalents do not correspond to the time pattern of learning and the methods used to produce them may not represent equal learning. Evaluation experts have often criticized use of grade-equivalents. To quote from "A Prototype Guide to Measuring Achievement Level and Program Impact on Achievement in Bilingual Projects" (Horst et al., 1980):

> They are based on the mistaken belief that a gain in test scores of one or more months for each month of instruction represent[s] good progress. This is not true. Grade-equivalent scores provide an illusion of simplicity but, in fact, they are almost impossible to interpret, even for specialists in test construction. Grade-equivalent scores should never be used by anyone for any purpose whatsoever. (emphasis added)

However, there is disagreement among testing experts whether grade-equivalents are totally unacceptable for measuring student achievement. One study (Olesini, 1971) which would have been accepted except for its use of grade-equivalents has been separately identified.

Outline of This Report

The remaining chapters of the report discuss the application of the accepted standards of scientific research (see Campbell and Stanley, 1963, for a partial list) to the literature on the effectiveness of bilingual education. Chapter 2 discusses the studies we found to be applicable to our questions, focusing on what limits are imposed in generalizing from these studies. Chapter 3 discusses why we found a number of studies--including several studies widely cited by proponents of transitional bilingual education as evidence of the effectiveness of such education--not to be applicable to our concerns. Chapter 4 presents our conclusions.

NOTES

1. Because a literature review is a secondary analysis, and the level of detail provided by the authors varies, we are limited as to what can be learned from the studies reviewed. These limitations were taken into account when we drew conclusions from the literature.

2. The methodological criteria we consider are generally accepted principles of social science research, such as those discussed by Campbell and Stanley (1963).

3. Any review is based on professional judgment, although this fact is often not stated. People who disagree with our findings may argue that we applied arbitrary criteria and personal judgment. Therefore we have carefully explained our methodology so that readers can assess for themselves how we made our scientific evaluation. Our effort to explain this process makes the paper long.

4. Studies that used a norm-referenced design include Rimm, 1980a; Young, 1980; Stern, 1975; Corpus Christi, 1980a, 1980b; JDRP, 1977b; St. John Valley, 1980; AIR, 1975e; Ames and Bicks, 1978; Arce, 1979; Fairfax County, 1980.

5. The following are examples of studies that were given no further consideration because they used a posttest-only design: South San Francisco, 1979; Elligett, 1980; AIR, 1975e, JDRP, 1977b; St. John Valley, 1980; Del Buono, 1971.

6. The following studies looked at "pre-post" gain with no control group: Guerrero, 1980; Liberty Union, 1980; Birmingham, n.d.; Valencia, 1971, Smith, 1978; Cahill and Foley, 1973; Ghini, 1979; B. Cohen, 1971; Stern, 1975; Price, 1978; South San Francisco, 1979; Charlotte-Mecklenberg, 1980; Arce and Sosa, 1975.

7. Studies that used grade-equivalent (GE) scores include Educational Management Services, 1976; Young, 1980; External Evaluators, 1979; Giles, 1971; Trevino, 1970; Kaufman, 1968; Olesini, 1971; Leyba, 1978; Hanson, 1980; Elligett, 1980; South San Francisco, 1979; Charlotte Mecklenberg, 1980; JDRP, 1977b; St. John Valley, 1980; AIR, 1975e, Arce, 1979.

8. The glossary in appendix A gives more detail on how these variables affect learning in bilingual education.

9. We should also note that equating the treatment and comparison groups for IQ or initial achievement probably does not account for this type of ability. We encountered no evidence in our literature review suggesting that any of the widely used tests are perfectly correlated with innate language ability. Therefore, to the extent the tests are unrelated to innate language ability, efforts to control statistically for IQ and pretest will fail.

CHAPTER 2

STUDIES APPLICABLE TO THE ISSUE

After gathering the studies of bilingual programs, we read and analyzed each study to see if there were reasons why the study was not applicable to the issue of interest. Chapter 3 discusses those studies we found not applicable and explains our reasons for rejecting them. This chapter discusses the 28 studies we found applicable.

The studies we found applicable exhibit the following general characteristics:

- Five studies were true experiments with random assignment (Plante, 1976; Pena-Hughes and Solis, 1980, 1981; Covey, 1973; Kaufman, 1968; Huzar, 1973) and one study used random assignment in three of four schools studied (Lum, 1971).
- In studies with nonrandom assignment, something was done--generally analysis of covariance--to adjust for possible differences between groups, either statistically or by matching.
- Two studies seem to have overcome most of the problems of nonrandom assignment by very thorough and clever study designs (McConnell, 1980a, 1980b; Lambert and Tucker, 1972).
- By definition, none of the studies relied exclusively on grade-equivalent scores or on the norm-referenced design.

These 28 studies have reported program outcomes about which we can be reasonably confident. However, our interpretation may differ from the conclusion presented by the authors. Consider, for example, the case of Balasubramonian et al. (1973). These authors concluded that bilingual education was very successful because, in comparison to students in the English as a second language program, students in the transitional bilingual program did not perform worse in English (L2), while they were able to to improve their skills in their first language (L1). According to our criteria specified in chapter 1, however, this study failed to demonstrate the effectiveness of transitional bilingual education, since neither English performance nor performance in nonlanguage subjects was found to be superior in the students in the English as a second language program.

Social science research is rarely, if ever, completely free of methodological problems. Those studies that we accepted as meeting our minimal methodological criteria are not without problems. The discussion of the studies' problems will therefore establish the upper limits of confidence we can place in the authors' conclusions.

Table 2-1 summarizes the important characteristics of the various types of instructional programs used to meet the special needs of language-minority students. The third column, special curriculum, refers to whether the curriculum is organized differently from the way the curriculum is organized in an ordinary monolingual program. For example, TBE does not have a special curriculum because it uses a regular curriculum in two languages, whereas the immersion method does involve a special curriculum.

TABLE 2-1. TYPES OF INSTRUCTIONAL PROGRAMS

	Language Spoken by Teacher	Language Spoken by Student	Special Curriculum
Submersion	L2	L2	No
Immersion	L2	L1 and/or L2	Yes
Transitional bilingual education	L1 gradually replaced by L2	L1 gradually replaced by L2	No
English as a second language	L2	L2	Yes

L1 = the child's first or home language.
L2 = the child's second language, the language of the school.

The rest of this chapter will discuss the 28 studies we accepted. Studies of TBE will be discussed first, followed by immersion studies. ESL is discussed in studies when it is compared to either transitional bilingual education or immersion. For each study, we will provide a description, point out the methodological strengths, and then discuss the findings.

TRANSITIONAL BILINGUAL EDUCATION

(TBE)

Name of Study : An Analytical Study of Secondary Freshman Bilingual Education and Its Effect on Academic Achievement and Attitude of Mexican-Americans

Author and Date : Covey (1973, 1981)

Location : Phoenix, Arizona

Treatment Group* : 100 Mexican-Americans in bilingual education program

Comparison/Control Group : 100 Mexican-Americans in regular English curriculum

Duration : 9 months

Ages : Grade 9

Type of Program : Transitional bilingual education

Description

Covey's (1973) study of a bilingual program at the largest high school in Phoenix is important for two reasons: First, it is one of the few studies of program effectiveness at the secondary level. Second, it is one of only six studies using random assignment. Unfortunately, Covey's study is very uninformative as to the nature of the program. The school year 1970-71 was the initial year of the ninth-grade bilingual program, but no further details were given as to program structure. Mexican-American students qualified for the sample if they met at least one of the following criteria: (1) they demonstrated a limited ability to speak English, (2) they came from a bilingual home environment, (3) they manifested a reading deficiency, or (4) they showed a deficiency in English and mathematics.

The school randomly divided 200 eligible students between the program and regular classes. The tests used were carefully selected and the following results were obtained at the end of the year:

- On the Iowa Test of Educational Development, students in the program outperformed the control group on the subtest "Correctness and Appropriateness of Expression." There was no difference in the performance of the two groups on "Ability to Do Quantitative Thinking" (math).
- The experimental group scored higher on the Stanford Diagnostic Reading Test.

---

* These brief descriptors sometimes oversimplify complex design features and appear here to give the reader a rough idea of the scope of each study.

Strengths

The study used random assignment to treatment and control conditions. A fairly large sample size was involved with a variety of outcomes being measured.

Discussion

Although Covey demonstrates a statistically significant improvement in test scores, the program was not universally successful. Only 10 to 15 percent of the program participants achieved a sufficiently high level of English proficiency to be mainstreamed (Covey, 1981).

Problems in interpreting Covey's results stem from the nature of the program. Apparently, the Phoenix program was an individualized diagnostic/prescriptive program. Participants spent 2-1/2 hours a day on reading, math, and English. Spanish was spoken as needed. Including aides, the pupil-instructor ratio was about 8 to 1. Very high parent participation was attained with considerable use of volunteer parents in the classroom. Thus, when parent volunteers are considered, the pupil-teacher ratio was even lower (Covey, 1981).

Hence, there are three competing explanations as to why the program worked: (1) individualized instruction, (2) low pupil-teacher ratio, and (3) bilingual instruction.

Name of Study : Effectiveness of Individualized Bilingual Instruction for Migrant Students

Author and Date : McConnell (1980a, 1980b)

Location : Washington State and Texas

Treatment Group : 630 migrant children with Spanish as the primary language in a regular English curriculum

Comparison/Control Group : 390 migrant children with Spanish as the primary language

Ages : 5-9 years old (grades K-3)

Type of Program : Transitional bilingual education

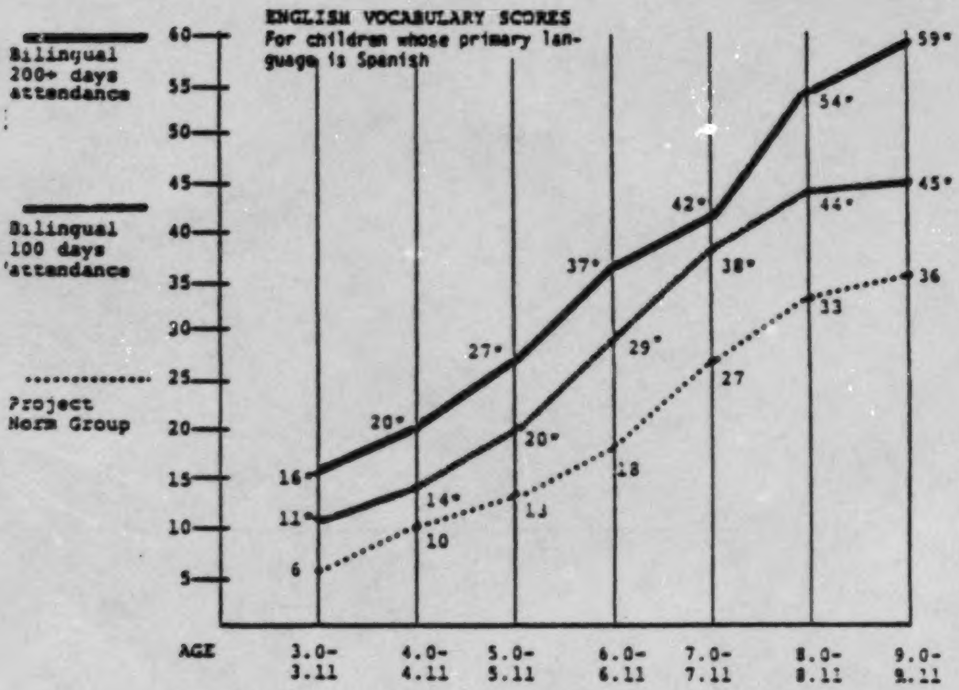
Description

McConnell has written two reports examining an individualized bilingual instructional (IBI) program for migrants. The IBI program for Spanish-speaking migrant children had base schools at both the winter home in Texas and the summer work location in Washington State. The Texas site was located in a community where there was an emphasis on preserving the Spanish heritage while the Washington community placed more emphasis on development of English skills. Students in the community emphasizing English did better on English measures while students from the community emphasizing Spanish did better in Spanish. Some of the program teachers went along on the migration to provide educational continuity between the two home camps. Schooling was available 12 months a year. A program of individualized instruction was used for math, reading, English, and Spanish. The program was for preschool through grade 3.

The method of analysis differs considerably between the two reports. In McConnell (1980a) the comparison group was formed by using the program students' test scores at entry into the program as age-adjusted pretests. Students of varying ages enter the program at different times and are pretested. By accumulating these scores over the years for each age level, the program generates an age specific comparison group for which there is little issue of selection bias since the comparison group was the group selected.

McConnell's figure 4 is reproduced (see table 2-2) to help the reader get a better grasp of the method. Each asterisk in the figure indicates the results of a t-test comparing the "project norm group" to an age-specific posttest score. Although such a use of multiple t-tests is not the best way to analyze the data, it is obvious from the figure that the results are robust and there is little use to worry over this point.

TABLE 2-2. McCONNELL FIGURE 4



*The superiority of this score over that of the project norm group of the same age is statistically significant beyond the .001 level.

Detailed test score analysis is shown in Table 6 in the Technical Appendix.

FIGURE 4. ENGLISH VOCABULARY SCORES ON FORM A, PEABODY PICTURE VOCABULARY TEST, OF CHILDREN WHOSE PRIMARY LANGUAGE IS SPANISH, BY AGE AND ATTENDANCE GROUP COMPARED TO THE PROJECT NORM GROUP.

TO SUMMARIZE THE FINDINGS IN FIGURE 4:

1. COMPARISON TO THE AVERAGE SCORES IN ENGLISH VOCABULARY OF THE PROJECT NORM GROUP SHOWS SIGNIFICANT SUPERIORITY FOR CHILDREN IN THE IBI BILINGUAL PROGRAM AT EVERY AGE LEVEL.
2. THE SUPERIORITY OF CHILDREN WHO ATTENDED FOR 200 OR MORE DAYS IS STATISTICALLY SIGNIFICANT BEYOND THE .001 LEVEL (E.G., THE POSSIBILITY THAT THIS MUCH DIFFERENCE WOULD OCCUR BY CHANCE IS LESS THAN ONE IN 1,000).
3. CHILDREN AFTER 200 OR MORE DAYS ATTENDANCE ARE MARKEDLY SUPERIOR TO THOSE TESTED AFTER ONLY 100 DAYS ATTENDANCE, INDICATING THAT THE GAINS IN ENGLISH ARE PROGRESSIVE THE LONGER THE PERIOD OF ATTENDANCE.

Based on this analytical model, McConnell found the following results:

- Improved Spanish performance in both Spanish- and English-dominant students.
- Improved vocabulary (English) in Spanish-dominant children.
- Improved math scores.
- Improved reading performance in Spanish-dominant children with 200 or more days in the program.

In a second analysis McConnell looked at the performance of only the children in the mobile component of the program. In the entire project, some students participated only in the Texas school, some participated only in the Washington school, and some participated in both places. Instruction was also provided while on the road, where the project provided staff to accompany the migrant caravan (the mobile component).

Apparently the mobile component group of students was too small to generate enough data to be able to apply the design described above. An additional comparison group was formed by taking Spanish-dominant migrant students from a neighboring school, grades K-3. The program students did significantly better in English vocabulary, math, and reading (English). The more time participants spent in the program the better they did in these three areas, also.

Strengths

The study revealed a well-designed longitudinal analysis with a large sample size and a variety of measures and comparison groups. One comparison method involved using the student as his or her own control. As children of different ages entered the program, their pretest scores became the comparison scores at posttest for children who were younger by the pre-post interval.

Discussion

McConnell's (1980a) study has the same problem of competing explanations as does Covey's (1973) study. There are several alternative explanations for the treatment group's improvement in performance:

1. Individualized Instruction. Many educators have held that individualized instruction is far superior to the standard classroom setting. The students in McConnell's study had individualized instruction in very small instructional groups (with a student-instructor ratio of 10 to 1).

2. A Coordinated Education. A major problem in the education of migrant children is that moving from school to school disrupts education. Their new schools are very unlikely to be at the point in the curriculum where the students left off when they left their old school. This results in serious gaps in education, often leaving children without the prerequisites for mastering later skills and therefore leading to very poor

educational performance. The individualized bilingual instruction program provided a continuous curriculum. For the first time in their school experience, these students were able to experience the normal school sequence.

3. Continuous Education Was Available. The number of instructional days received was not reported. However, having services available 12 months of the year made it possible for children in the program to receive significantly more days of instruction during each grade than the comparison students did. One of the best established principles in educational research is that time on task (in this case, days of exposure to schooling) is related to achievement.

Analysis of the mobile component revealed two design features that raise serious question about the results. First, the matching procedure was incomplete. Community differences were not controlled for in any of the analyses. More important, IBI students could have begun the program as early as age 3. Since no mention is made of a preschool program in the comparison community, the superiority of the IBI students could be attributed to the 2 to 3 additional years of formula instruction.

In the second report, McConnell (1980b), the comparison group was formed by combining the program participants' entry scores with the scores from the nonequivalent comparison group for the neighboring school. Unfortunately, this contamination of the participants' entry scores with the uncontrolled nonrandomly selected comparison group makes any analysis using the combined group doubtful.

The method of analysis is to compare test scores for five categories of length of attendance in the IBI program (zero to 3+ years) by age-standardized scores. Although the results consistently favored the program by showing higher scores as length of attendance increased, the analysis is suspect for two reasons. First, as we noted above, the zero-attendance group was contaminated by the inclusion of a nonrandomly selected comparison group from a neighboring school. Second, since each participant was tested twice, one in the zero-attendance group and again in one of the four length-of-attendance categories, the scores were not independent. This is a violation of one of the assumptions underlying the F-test used in the analysis.

Although McConnell (1980b) has problems with a confounded treatment group, violation of the assumptions of the tests, and use of a nonrandom comparison group with no adjustments made for preexisting differences, we conclude the results presented in McConnell (1980b) can be accepted for the following reasons:

- They are fully consistent with the results from McConnell (1980a).
- They are internally consistent across 3 years of data.
- They are internally consistent within the order of the means for the participant only group through four categories of length-of-program exposure (excluding the contaminated entry scores).

In short, McConnell's results show the findings are robust enough to hold up, even under the methodological problems introduced by the analysis plan in the second report (McConnell, 1980b).

Name of Study : A Study of the Effectiveness of the Connecticut "Pairing" Model of Bilingual-Bicultural Education

Author and Date : Plante (1976)

Location : New Haven, Connecticut

Treatment Group : 45 Spanish-dominant children

Comparison/Control Group : 27 Spanish-dominant children in a regular English curriculum

Duration : 2 years

Ages : Grades 1 and 2

Type of Program : Transitional bilingual education

Description

Plante (1976) reports on a well-designed study of New Haven's bilingual program. All Spanish-surnamed pupils completing kindergarten and first grade in an attendance area serving a large percentage of children from low-income families were given the Inter-America Test to identify their language dominance. Those children in grades 1 and 2 who were identified as being Spanish-dominant were randomly assigned to the bilingual program or to a control group. Thus, methodologically, the study was a true experiment.

No bilingual-bicultural instruction had been available to these children prior to the implementation of the study. The type of bilingual instruction investigated involved the "pairing" model in which one native Spanish-speaking teacher taught basic skills in Spanish and an English-speaking teacher taught speaking, reading, and writing in English.

The students were tested again after 2 years. In both grades there were large and significant differences in L1 reading favoring the bilingual program. In English, a significant difference favoring the bilingual program was found at grade 2 but not at grade 3 and not over both grades combined.

Strengths

The study used random selection and, thus, was a true experiment. Plante also collected data over a 2-year period.

Discussion

Plante was not justified in stating that "it seems significant to point out that the experimental pupils in the second grade and third grade, as well as when a total group analysis is made, exceed the English reading achievement of the control group." The only significant difference favoring TBE was found in one grade. TBE was found no different from the regular English curriculum in the other grade or for both grades combined.

Plante presents results from the Metropolitan Achievement Test in grade-equivalents. He provides no tests of the significance of the differences between the experimental and control groups, although all comparisons favored the bilingual group.

Unfortunately, Plante's carefully designed study is ultimately undone by the drastic difference between the experimental students and control group students in retention rates. All the experimental students except one progressed through 2 years of school in the 2 years of the study. Only half of the controls did so; the other half were retained. Thus, half of the control group had 1 year of school twice in the 2 years of the study. The minimal differences in test scores between the two groups indicate that the difference in failure rate was due to the fact that bilingual and regular classroom teachers applied very different promotion criteria to their respective students. The experimental and control groups clearly were treated very differently in ways that did not relate to language treatment. Consequently, we can conclude that Plante's study suggests a program effect but ultimately fails to prove that the language treatment was effective.

Name of Study : The Effects of Program Models on Language Acquisition by Spanish Speaking Children

Author and Date : Legarreta (1979)

Location : A large West Coast city never specified

Treatment Group : 80 monolingual Spanish-speaking children

Comparison/Control Group : Subsets of the treatment groups were compared to one another

Duration : 6 months

Ages : Kindergarten

Type of Program : Submersion, English as a second language, transitional bilingual education

Description

Legarreta (1979) tested kindergarten children drawn from existing programs. Teacher judgment, a language use questionnaire, and pretest scores were used to limit the study to students who were "essentially monolingual." Five instructional models were compared: (1) submersion, (2) ESL, (3) bilingual—concurrent translation, no ESL; (4) bilingual—balanced bilingual education, no ESL, and (5) bilingual—concurrent translation with ESL. The bilingual programs using concurrent translation were programs in which English-taught subject matter was immediately translated into Spanish. The bilingual program referred to by Legarreta as a "balanced bilingual" program, using 50 percent Spanish and 50 percent English, can also be characterized as an alternate immersion program.

Legarreta found bilingual programs superior to traditional submersion, and bilingual education programs with ESL superior to programs without ESL. Since one traditional program had ESL and two of the three bilingual programs were without ESL, it is impossible to draw any conclusions from Legarreta's data as to the relative importance of ESL and bilingual instruction. These effects are never separated.

However, Legarreta found that children in the balanced bilingual education class, who received less English instruction than the other two bilingual education classes, outperformed those groups on two measures of English skills. The unbalanced treatments received English instruction approximately 72 percent of the time.

Strengths

Legarreta's analysis is complex. Students were not randomly placed into the five types of treatment. However, analysis of covariance was used to adjust for preexisting differences. The study was restricted to monolingual, Spanish-speaking children, the type of children pertinent to transitional bilingual education theory.

Discussion

Legarreta found that the more instruction given in the native language, the better the performance in L1. But the next finding is not so obvious. The balanced bilinguals in the alternate immersion classroom were also found to outperform the unbalanced bilingual groups on two measures of English skills. In the case of Spanish, more instruction led to better performance. In the case of English, more instruction led to poorer performance.

At first glance, the language facilitation effect appears to provide a possible explanation for the result that more instruction in L2 led to poorer L2 performance.* The language transference theory is generally applied to reading, however, and Legarreta's subjects were all kindergarten children tested on oral language proficiency. The literature seems to differentiate oral and reading language skills. What applies to one does not necessarily apply to the other. Furthermore, to argue that the transference hypothesis applies to Legarreta's data contradicts other established principles of oral second-language learning. Oral L2 skills are best learned by young (ages 6 to 7 or younger) children—precisely the ages of Legarreta's students—and L2 mastery in young children is a direct function of practice.

There are alternative explanations for Legarreta's results. The seven students (one class) in the "balanced bilingual" program who outperformed students who spent more of their day in English could have had an unusually effective teacher. Such a small sample is sensitive to teacher differences. Exceptional teachers will get exceptional results in either language.

Another alternative explanation may be that Legarreta did not test the program effects she thinks she tested. Two of the three bilingual groups employed "concurrent translation"—that is, the teacher would state a concept in one language and then immediately restate it in the other language. Legarreta points out that this method may be less successful than alternate immersion (the method used with the balanced bilingual group) because students tune out the English, knowing it will be followed by a Spanish version which they understand.

Berke (1980), de Kanter (1980), and Cummins (1981) suggest that concurrent translation is not an effective instructional strategy. Berke argues that linguistic theory clearly implies that concurrent translation leads to linguistic confusion. The most effective learning of a second language occurs when the setting in which the second language is used is clearly differentiated from that in which the first language is spoken, even to the point of having different teachers speak the two languages. Therefore, Legarreta's concurrent translation (unbalanced bilingual) group may have performed more poorly than did the balanced bilinguals because of linguistic confusion inherent in the concurrent translation method.

---

* Briefly, the facilitation effect, or language transference theory, proposes that there are formal reading skills which transfer from one language to another. Therefore, it is advisable to teach reading first in L1 and to rely on transference to help with mastery of L2.

If Berke's, de Kanter's, and Cummins' arguments are correct, the implication of the study is not clear. The difference between the concurrent translation and alternate immersion groups is exactly the same as the difference between unbalanced and balanced bilingual instruction.* Legarreta interprets the results as a test of the latter, but linguistic theory suggests that the former interpretation may be more appropriate. In either case, the interpretations of Legarreta's results are not clear. Her conclusion that balanced bilingual education "clearly is most facilitative of acquisition of English as well as maintenance of Spanish" is questionable.

We have discussed Legarreta's study at some length for several reasons. As one of the best statistical analyses in the literature, it is an important study. However, the study also shows that good statistics are not enough. Numbers must be interpreted. In a field as complex and as fluid as bilingual education, interpretation is difficult.

---

* Of Legarreta's four significant tests, three involve comparing one or both of the concurrent translation programs with the alternate immersion program. This practice makes it impossible to say what her tests measured. For example, two of the tests pit balanced bilingual instruction against unbalanced bilingual instruction.

Name of Study : Bilingual Education Program

Author and Date : AIR (1975b)

Location : Corpus Christi, Texas

Treatment Group : 269 Title VII bilingual education students

Comparison/Control Group : 124 non-Title VII students

Duration : 9 months

Ages : Grades K and 1

Type of Program : Transitional bilingual education

Description

Although the report discusses secondary data from 4 years, apparently only in 1 year for grades K and 1 was any adjustment made for preexisting differences between the two groups. The authors refer to adjusted posttest means in their tables, although the text gives no discussion of how the adjustment was made. In kindergarten there were no differences between the treatment and control groups on five English tests. In the first grade, performance by the TBE program students was superior to that of the comparison group on two Spanish tests and on two English tests. Scores on two more English tests were reported to be significantly higher for the program students, but these results were not adjusted and were given in grade-equivalents.

In addition, first- and second-grade students were categorized by the number of prior years they had spent in the bilingual program. A significant (analysis unspecified) result showing increasing improvement in English performance with increasing length of time in the program was reported. Although the analyses described in the preceding paragraph were also conducted for a second cohort (without adjustment), the table showing increased effects over time was not replicated.

Strengths

The Corpus Christi sample was moderately large. The authors made an adjustment for possible preexisting differences resulting from nonrandom selection, presumably by analysis of covariance. They also made use of more than one outcome measure.

Discussion

A complete assessment of the generalizability of the study is made difficult by the lack of detail provided in AIR's (1975b) secondhand account. Apparently the student population of the program schools is largely Hispanic, for the authors note that no screening for eligibility was done. All participants were also volunteered by parents, a situation that cannot be fully compensated for by the statistical adjustment.

Generalizability is also limited by the study's coverage of only two grades (K and 1) in one school district.

Name of Study : Will Instruction in Reading Spanish Affect Ability in Reading English?

Author and Date : Kaufman (1968)

Location : New York City

Treatment Group : School A: 48 Spanish-speaking pupils receiving 4 sessions of Spanish a week; school B: 27 Spanish-speaking pupils receiving 3 sessions of Spanish a week

Comparison/Control Group : School A: 37 Spanish-speaking pupils in regular English curriculum; school B: 27 Spanish-speaking pupils in regular English curriculum

Duration : School A: 2 years; school B: 1 year

Ages : Junior high

Type of Program : Transitional bilingual education

Description

Kaufman (1968) examined the effect of bilingual education in Spanish and English in two junior high schools in New York City. Bilingual education at this time was just beginning to receive national attention. Spanish instruction was given to the experimental group four times a week for 45 minutes per session at school A and three times a week for 45 minutes per session at school B. Experimental and control groups received equivalent instruction in English. There were no significant differences between the experimental and control groups in school A. In school B, two of six tests significantly favored the bilingual group, leading Kaufman to conclude, "These findings suggest that there was some evidence of positive transfer of learning from instruction in reading Spanish to reading ability in English at school B." Altogether, Kaufman found a significant difference on two of nine tests he performed.

Strengths

Kaufman used random assignment in his study making it one of the six true experiments in the literature. He also used analysis of covariance to control for preexisting differences in verbal IQ, nonverbal IQ, age, and English pretest score.

Discussion

By analyzing the two schools separately, Kaufman partially lost the advantages of random assignment because school differences are confounded with the treatment. If the bilingual students at school B had an unusually effective teacher, Kaufman's results would be expected. Kaufman should have combined the data from the two schools in his analysis and tested for school and teacher effects. We can only approximate such tests using data from the article. In school B, the difference between the bilingual and control means was 0.39. That difference is reduced to 0.18 for the combined weighted means of both schools A and B. We cannot say whether this lower difference is significant because sample size is increased. The large reduction in the mean suggests that the difference Kaufman found between the bilingual and control groups in school B was due to something other than the effect of the program. On the other hand, the school B results could be indicative of the program effect while school A results could be due to an unusually poor teacher. The point is that something other than a program effect was at work, so caution must be exercised in drawing conclusions from the study.

Kaufman had longitudinal data measuring word meaning, paragraph meaning, and total score means for English reading. Since students in school A were posttested three times and students in school B were posttested twice for these 3 measures, Kaufman could have examined the hypothesis that, if the program had an effect, the difference between the bilingual and control groups would increase over time. We have computed these differences for Kaufman's three measures:

TABLE 2-3. DIFFERENCES IN TEST SCORE MEANS (KAUFMAN, 1968)

	School A			School B			N	
Test Point	Word Meaning	Paragraph Meaning	Total	Word Meaning	Paragraph Meaning	Total	A	B
1	-.0195	.1169	.903	.3933	.2772	.3422	78	49
2	.1143	.0328	.0747	.0290	.3032	.1430	72	45
3	.2152	.1128	.1542	NA	NA	NA	50	NA

There is no improvement over time for school A when the bilingual and control groups are compared. School B shows improvement in the bilingual class on the first posttest for word meaning. The total mean shows improvement but is really again measuring the effect of word meaning scores. The second posttest showed no significant improvement.

Interestingly, school A had the most Spanish instruction and showed no English reading improvement. However, school B had one period less Spanish instruction and showed significant gains. This, again, suggests that something other than bilingual instruction accounted for the achievement differences in school B.

Name of Study : An Evaluation of the Effectiveness of Selected Experimental Bilingual Education Programs in Connecticut

Author and Date : Zirkel (1972)

Location : Connecticut (Bridgeport, Hartford, New Britain, New London)

Treatment Group : Grade 1: 73 Puerto Rican children; grades 2 and 3: 53 Puerto Rican children

Comparison/Control Group : Grade 1: 73 Puerto Rican children in a regular English curriculum; grades 2 and 3: 76 Puerto Rican children in a regular English curriculum with ESL

Duration : 1 year

Ages : Grades 1-3

Type of Program : Transitional bilingual education

Description

Zirkel (1972) looked at three variations of bilingual programs in four towns. The author points out that in some cases what was termed a "bilingual" class in one location was indistinguishable from what had been indicated as a "control" class in another area. The author characterized these variations as (1) the bilingual education model and (2) the quasi-bilingual model. The children involved in the study were "economically disadvantaged" Puerto Ricans. The use of L1 in the classroom varied considerably in the schools included in the study, ranging from 10 to 150+ minutes daily.

Adjusted posttest means in both English and Spanish reading were analyzed by five tests. There was only one significant difference favoring the bilingual program in English (and one in Spanish for the same group of students). This difference occurred in grades 2 and 3, which had the most English instruction. The analysis of quasi-bilingual models disclosed nonsignificant differences between the experimental and control groups in achievement outcomes.

Strengths

To partially control for the effects of nonrandom selection, Zirkel matched the experimental and comparison groups for age, sex, and socioeconomic status by eliminating students from the study until group level matching was attained. Analysis of covariance controlled for pretest and nonverbal IQ.

Discussion

Bilingual teachers in grades 2 and 3 were more qualified than first-grade teachers. Therefore, teacher differences probably had a significant impact on the higher level of student achievement in grades 2 and 3.

Name of Study : An Evaluation of Title VII Bilingual/Bicultural Program, 1977-1978 School Year, Final Report

Authors and Date : Ames and Bicks (1978)

Location : Brooklyn, New York

Treatment Group : 212 Spanish-speaking and Creole-French-speaking children in bilingual and ESL classes

Comparison/Control Group: 457 Spanish-speaking and Creole-French-speaking children in regular English curriculum with some ESL pullout

Duration : 9 months

Ages : Grades 1-9

Type of Program : Transitional bilingual education, English as a second language

Description

Ames and Bicks (1978) present an interesting study of a bilingual program in New York City. Participants in the program were Spanish speakers and Creole French speakers who scored below the 20th percentile on the English Language Assessment Battery developed by the New York City Board of Education. The comparison group was composed of students with non-English-language backgrounds in schools where the bilingual population was not large enough to justify a full bilingual program, but where the same English-language instruction was given on a pullout basis as was done in the schools with bilingual programs. The bilingual and pullout groups of students received 3 to 5 hours of English instruction a week. It appears the ESL group had more English instruction than the other two groups.

The size of the bilingual and ESL classes was less than 25 students. All teachers were fluent bilinguals. Some of the teachers used individualized instruction, while others chose "traditional" styles of teaching. A comparison of the bilingual versus ESL versus pullout groups in English reading showed significance. Although the bilingual and pullout groups had equal gains, the ESL group had the largest gain.

The authors conclude:

> Since the time spent in intensive study in English was the same in both the bilingual and pullout groups, it was understandable that achievement in reading English was not significantly different. In math, however,...the results...indicated that those students who received instruction in the native language achieved higher scores in math than those whose instruction was given in English.

Strengths

Analysis of covariance was used by Ames and Bicks to adjust for preexisting differences due to nonrandom assignment. The method of selecting the comparison group was very well devised and inspires more confidence in the equivalence of the treatment and control groups than is usually the case.

Discussion

The ESL group is never full described in the report. We are never sure how much English instruction they received. However, on the reading test the ESL students showed gains 50 percent higher than either the bilingual or pullout students. The authors present an insignificant F-test for bilingual versus pullout, but, unfortunately, no separate test of ESL versus either of the other two program types was given. In math the pullout group gained 49 percent of the bilingual group's gain, but the ESL group achieved 80 percent of the bilingual group's gain. Again, no paired tests involving ESL were given. Because no tests of ESL other than the overall F-test were given, and because we are unsure whether the authors used adjusted or raw means, it is unclear whether the ESL program may have been equal or superior to the bilingual program.

Another problem with the analysis is that the bilingual versus pullout test was apparently a two-variable analysis of covariance done after the three-variable test. This is not the preferred way of testing differences between pairs of means after finding a significant F-test in the omnibus test.

The study may also be confounded by teacher differences, as some teachers employed individualized instruction and others employed traditional methods.

Name of Study : Iloilo I and Rizal

Authors and Date : Ramos, Aguilar, and Sibayan (1967)

Location : The Philippines

Treatment Group : Iloilo: 232 students instructed in Hiligaynon and then English; Rizal: data unavailable

Comparison/Control Group : Iloilo: 301 students instructed in English; Rizal: data unavailable

Duration : Iloilo: 6 years; Rizal: 5 years

Ages : Iloilo: Grades 1-6; Rizal: Grades 1-5

Type of Program : Transitional bilingual education/immersion

Engle (1975) thoroughly described a 20-year study of instruction in the first language in the Philippines, which we quote here. We have divided her description into sections for clarity.

> Iloilo I and Rizal (Ramos et al., 1967). The Philippine studies referred to by these place names represent two carefully designed studies executed between 1948 and 1967. In some respects, the results conflict with each other, and one can begin to understand some of the factors only by examining the differences between the two [and] then integrating the findings. The language situation in the Philippines is complex; there are many vernacular languages. Tagalog (or Pilipino) is the national language, and the government would like to encourage English in the elementary schools as an additional national language.

Description: (Iloilo I)

> The original Iloilo experiment was conducted in an area of the Philippines that speaks Hiligaynon as its mother tongue. The Iloilo I study was designed to analyze the effects of initial instruction in the vernacular on the eventual learning of the curriculum in English.
>
> The study was statistically well designed. The experimental group received instruction in the vernacular for grades 1 and 2, and in English in grades 3 through 6. The shift was abrupt. The controls were given all instruction in English from grades 1 through 6. Instructional materials were identical throughout, with the exception that the first and second grade experimental materials were translated from English into Hiligaynon. Assessment of abilities and achievement occurred before grade 1, and

after each grade through grade 6. Assessment in grades 1 and 2 was in the vernacular for the experimental group, and in English for the controls. All of the assessment was in English for the 3rd through 6th grades. Tests of reading, arithmetic problems, understanding social studies, and learning language skills were given at each grade.

At the beginning of the project the sample included 1,164 controls and 758 experimentals. When the experimentals were matched with controls, the sample size decreased to 188 experimentals and 189 controls. Unfortunately, the attrition rate was very high; by the end of the 6th grade, only 28 percent of the sample remained. Reports of the actual numbers differ. The report of the 6th grade evaluation indicates that 232 E's and 301 C's remained in the study. Of these, 82 were matched.

The methods employed for second language teaching were different from the standard methods. The new method was based on language patterning and drills, emphasizing both structure and sound relationships. Teachers were given training in teaching in Hiligaynon and in teaching English as a second language.

At the end of the first year of the study, the experimental group was significantly higher on reading (in the language of instruction) and social studies. The differences in arithmetic were not significant. At the end of the fourth year (two years of instruction in English for the experimental group, four for the control), a nonsignificant superiority in the control group was found for reading and arithmetic, and a significant superiority was demonstrated for language. The experimental group had a slight superiority in social studies.

Strengths: Iloilo I

The sample consisted of fourteen elementary schools equated for SES, teacher quality, the principal's qualifications, and supervisors. Experimental teachers were generally of higher SES than controls. Children were further equated on the Philippines Mental Ability Test, chronological age, and school attendance.

Discussion: Iloilo I

The confusion with which the project is reported and quoted is exemplified in a comment by Venezky (1970) who reviewed the study. He reports that an independent investigation of the fourth-year results by the Director of Public Schools showed significant superior performance by the controls on all tests, including social studies. Venezky includes no references. He alludes to the "overenthusiasm" of the program's director for the native language approach as a confounding factor, and this observation of enthusiasm is echoed in the report by Ramos et al. (1967). Although one cannot discard a study because the director believes in it, one can suggest that the Hawthorne effect may be operating in a situation where one group is seen as more exciting and more significant than the other. No control group matched on "being studied" was included, nor did the investigators seem aware of this problem.

In the sixth year evaluation (with the reduced sample size), the experimentals were superior to the controls in social studies achievement tests and slightly higher in arithmetic and reading tests. The controls scored slightly higher in language.

A Personality Inventory was given at the end of grades 4, 5, and 6. Children in the experimental groups reported themselves significantly higher on one of the 4 or 5 dimensions of that test, though the dimension varied from year to year.

The results of this study were widely accepted in the Philippines; all children were then given instructions in their vernacular for the first two years of school, and in English for the remainder. We feel that such an adoption of policy was premature, particularly when the sample was so small, and when only one model had been attempted. Other variations, such as introducing the second language in the first grade as a language of instruction, were not mentioned.

The study suffers from a number of problems typical of many such studies. The tests were inadequately validated in English and then simply translated into the vernacular. Variables were not isolated; a new method of instruction was confounded with the basic hypothesis (differences in language of instruction). No control for the Hawthorne effect was made; the children could well have been achieving because they felt special. The two curricula in first and second grades were unequal since the English materials were published and polished, whereas the Hiligaynon lessons were on "rough dittos, often unclear." The level of knowledge of English on the part of the teachers is recognized as extremely low. The high drop-out rate suggests that the final sample is extremely select in terms of the factors which permit a child to stay in school—probably related to SES.

Even though severe criticisms can be raised methodologically about the interpretation of the study as a test of the reading transfer and mode-of-instruction hypotheses, it does indicate clearly that experimental children in this situation were not hampered in achievement. If, that is, they were able to stay in school for 6 years.

Description: Rizal

The Rizal study was designed to gather information as to the most appropriate time to introduce reading in English and English as a medium of instruction, questions that had not been answered by Iloilo I. Five groups were defined according to the children's grade level at the introduction of English for reading and as a medium of instruction. F. B. Davies of Hunter College, New York, served as consultant for these two studies.

Grade in which English is first used as a medium of instruction	Grade in which English reading begins	
	First	Second
First	Group 1	
Third	Group 2	Group 4
Fifth	Group 3	Group 5

Teachers received instruction on the teaching of English, and teaching various subject matters in English. They received no instruction in the teaching of, or with, the vernacular (in that location, Pilipino) and continued to use outdated material that had followed from the recommendations made after the first Iloilo experiment.

The time at which reading in English was introduced apparently made little difference on an English language reading test.

To test the hypothesis concerning the medium of instruction, three versions of all achievement tests were constructed: English, Pilipino, and bilingual. The results suggested that varying the medium of instruction did not have a large effect on basic skills. The only effects were on arithmetic and language scores. Those who had been introduced to English most recently scored highest on the English arithmetic test. Those who had used English as their medium of instruction the longest had the highest scores on the English language tests in sixth grade.

Strengths: Rizal

Schools were systematically selected, equated on significant variables, and carefully matched.

Discussion: Rizal

The authors argue that English competence is directly related to number of years studying English. One might therefore argue that if you want children to learn English, you should begin as soon as possible with English, since

> it does not retard the other language and the children will have learned English better. These results contradict the findings of Iloilo I. However, on closer inspection a third variable might explain the discrepancy.
>
> As noted previously, the teachers were not trained in the use of the vernacular in the Rizal study; all of their training was in the use of English. On the other hand, the majority of the training in the first study was in the use of the vernacular for instruction. Thus a combination of teacher training effects and Hawthorne effects might explain the difference between studies. Ramos et al. (1967) themselves suggest such an explanation. A future direction, then, could be an assessment of teachers' capabilities and the effects of teacher training on achievement in various areas. However, neither of the Philippine studies conclusively indicates whether introducing reading in the first language should be introduced as a language of instruction.

We conclude, as does Engle, that these studies do not clearly show consistent superiority of either transitional bilingual education or immersion. Therefore, we classify the studies as a single study showing no difference between bilingual education and alternatives.

Name of Study : An Investigation of the Effects of Background Characteristics and Special Language Service on the Reading Achievement and English Fluency of Bilingual Schools

Author and Date : Matthews (1979)

Location : Seattle, Washington

Treatment Group : 383 language-minority students in bilingual and ESL programs

Comparison/Control Group : 1,011 language-minority students in a regular English curriculum

Duration : 9 months

Ages : Grades 2, 4, 6, and 8

Type of Program : Transitional bilingual education, English as a second language

Description

Matthews (1979) analyzed the differences between all language-minority children in the Seattle special language service bilingual program and all language-minority children not in the program. Special language services included English as a second language and/or bilingual instruction. The largest number of children with a non-English-speaking background were Chinese, with Philippine, Korean, Spanish, Japanese, Samoan, Vietnamese, and other ethnic groups represented as well.

The exact nature of the program Matthews examined is unclear, being referred to as "Special language services," which was defined as including "English as a second language...and/or bilingual instruction in required subject matter: mathematics, science, health, social studies, language arts." However, the program was established under a Lau compliance agreement calling for TBE wherever practicable.

Matthews presents data on all language-minority students tested in grades 2, 4, 6, and 8 in the Seattle schools in the spring of 1979. The following control variables were used in the analysis: previous test scores from fall 1978 for grades 2, 6, and 8; the five-point Lau scale of English fluency; ethnic background; and free lunch (a proxy for family socioeconomic status). A theoretical model relating the control variables to each other and to achievement was developed. The components of the model were tested by partial gamma coefficients and X^2, a procedure somewhat similar to log linear modeling. About one-third of all bilingual students received bilingual services, with 56 percent of the students scoring in the lowest three stanines being served. Thus, approximately equal numbers of low scoring students were in the treatment and control conditions, which is very desirable from the perspective of the analysis.

The effect of TBE on achievement was tested by holding fluency, language background, income, and grade constant, forming 64 contrasts. There were six significant coefficients, all indicating lower achievement on the MAT total reading score (English) for students in the program.

In a second analysis Matthews examined gain from fall 1978 to spring 1979 (1 year of school) and found no difference between the proportions of students gaining against the norm in the treatment and control groups. Matthews states:

> ...served students tend to score lower than not served students even when background variables such a fluency and previous achievement are held constant. In addition, the served and not served students tend to improve their reading percentile scores in similar proportions (thus students who are not served progress in English reading skills as rapidly as those who are served).

Strengths

With a larger number of students, Matthews was able to carry out an analysis that has a number of advantages over the more usual approach using analysis of covariance. By using nonparametric statistics, the analysis avoids having to make a number of problematic assumptions about the nature of the underlying distribution and about the metric that are necessary in a parametric analyses. The one clear advantage parametric methods have, their greater power in detecting a difference, is not necessarily a problem here since the use of less powerful tests makes the analysis more conservative. Matthews may have overlooked a small true effect by using less powerful tests, but if so, the magnitude of the effect would be so small as to have little practical value.

Discussion

The generalizability of Matthews' findings is limited primarily by the restriction to one school district and by the relatively low proportion of Hispanic students in the school (only 9 percent of the language-minority students were Spanish speaking). An occasional problem is that, even with the large number of students, some of the 64 cells in the tests table contain so few students that the X^2 test may not be accurate (a point Matthews recognized).

It is worth repeating Matthews' final statement because it reminds us of the limits ultimately imposed on all the nonexperimental evaluations discussed here:

> Given the limitations of the current data base, it is impossible to determine with any confidence whether the results are a reflection of the effects of service or whether they merely reflect current implementation practices.

Name of Study : An Evaluation of Some Cognitive and Affective Aspects of a Spanish-English Bilingual Education Program

Author and Date : Skoczylas (1972)

Location : A community in northern California

Treatment Group : 25 Title VII students (Anglos and Mexican-Americans)

Comparison/Control Group : 22 Anglos and Mexican-Americans in a regular English curriculum

Duration : 9 months

Ages : Grade 1

Type of Program : Transitional bilingual education

Description

Skoczylas (1972) reports on a Title VII bilingual program for children in first grade. The bilingual education class and the comparison class came from two different schools in the same district in northern California. Each class contained native speakers of Spanish and native speakers of English. The experimental class was composed in part of the same children who had been in the bilingual kindergarten. Parents were given the option of enrolling their children in one program or another.

Approximately half of the teaching day in the bilingual program was devoted to activities conducted in each of the two languages. One teacher and two aides, all three bilingual, taught subject matter.

Skoczylas conducted a parent survey and extensive pretesting to see how the comparison and program groups might have differed on the relevant variables of age, IQ, home educational environment, school attendance, parents' educational background, language development, and sex. The two groups differed significantly on three of the background variables. The background variables were used as covariants in the analysis. This analysis is one of the best we encountered for taking into account the relevant variables.

Bilingual instruction did not lead to better English performance. Spanish performance of the program group was better, but math performance was worse. Skoczylas' study was limited to the first grade, so blanket generalizations to all grades are not justified.

Strengths

The author collected extensive background information from the parents, tested for initial equivalence of the groups, and used analysis of covariance to adjust for initial differences.

Discussion

Skoczylas (1972) found no difference between program and comparison first-grade students in Spanish listening comprehension and English skills. The program students performed significantly better in Spanish-speaking skills and significantly worse in math.

We paraphrase NIE's (1981) evaluation of the Skoczylas study, highlighting their criticisms:

Skoczylas' listening comprehension test (one of two tests of English used) consisted of 11 yes/no responses to questions about a brief paragraph. Thus, a child could be expected to get 4 to 7 correct responses by chance. On the Spanish version of this comprehension test, the mean pretest score for the control group was 2.2, well below what would be predicted even by chance response to the test. By posttest time at the end of the year, this raw mean climbed to 4.8, essentially moving into the range of chance response. (For the experimental group, raw group mean scores were 6.0 and 8.0 for pre- and posttest.) The experimental group was rated 48 percent Spanish-dominant or Spanish-monolingual; 23 percent of the members of the control group were in these categories. This, along with the home language usage data, seriously calls in question just how bilingual the children in the control group were. The analysis of covariance did not adjust for language use in the home.

The children in the experimental program were superior in the pretest to the children in the control group on three of the four measures of language use, including both measures of Spanish use and one measure of English use (the other English measure showed no significant difference). This difference in groups is important when one considers that the children in the experimental group had gone through a year of bilingual instruction in kindergarten, while the children in the other treatment had gone through regular kindergarten. This pretest is in some sense a commentary on the effects of the bilingual kindergarten program. It is in large part the superior performance of the children who had been in the bilingual kindergarten that forces the use of analysis of covariance in the study.

All the children in both groups took the math test in English. Half the children in the experimental bilingual program had had math instruction only in Spanish, with no prior math instruction in English. Thus, these children may have encountered the technical terminology of math concepts in English for the first time on the posttest.

There was no pretest in math, only a posttest. Nonetheless, the analysis of covariance had the effect of depressing the math scores of the children in the experimental bilingual program and increasing the scores of the children in the control group, because it was compensating for the superior language performance of the experimental children on the three language measures in the pretest. There are insufficient data presented to determine if the groups would continue to be significantly different if actual scores, rather than adjusted scores, had been used.

Name of Study	:	An Effectiveness Study of English as a Second Language (ESL) and Chinese Bilingual Methods
Author and Date	:	Lum (1971)
Location	:	San Francisco, California
Treatment Group	:	35 monolingual Chinese students in a bilingual class
Comparison/Control Group	:	20 monolingual Chinese students in an ESL class
Duration	:	9 months
Ages	:	Grade 1
Type of Program	:	Transitional bilingual education, English as a second language

Description

Lum (1971) compared a program of English as a second language with transitional bilingual education. The students were 55 monolingual Chinese-speaking first graders in San Francisco. Chinese language arts were taught in the TBE schools but not in the ESL schools. Although a little Chinese was used in the ESL program for content-area instruction, much more was spoken in the TBE schools. Use of Chinese in both schools declined over the school year, becoming very minimal by the end of the ESL school year. English language arts instruction was given through ESL procedures in both schools, but the ESL-only schools averaged 50 percent more time each day in ESL instruction.

All the teachers involved in the study were Chinese. The bilingual classes were team taught, so that one of the two teachers was expert in Chinese. The ESL classes had only one teacher per classroom.

Assignment to the treatments was rather complex. First, students were screened by self-report, teacher judgments, and the Hoffman Bilingual Scale to identify only monolingual Chinese-speaking students for the study. Students who lived in one area were randomly assigned to one ESL and two TBE schools. Students living in a second attendance area were all assigned to an ESL school. There were 35 bilingual and 20 ESL students, with two ESL and three TBE classes. In reviewing the assignment process Lum concluded that "subjects seemed matched through pretesting and randomization by area of residence."

Lum measured oral proficiency in English using ratings of tape-recorded responses to teacher-administered stimulus pictures. Apparently, the pictures and rating method were taken from standard procedures described in the literature, although Lum's wording implies they were modified.

Only one person did the scoring of the tapes. In general, free-response scoring systems are best done by multiple raters in order to achieve acceptable levels of reliability.

The students' responses to the pictures were rated on five scales: length of response, length of the five longest responses, number of different words used, structural complexity, and grammar. There were no differences between the groups on the last two items. On the first three measures, the English as a second language students significantly outperformed the transitional bilingual education students. Lum converted the scores to age norms and found that the TBE students were, in the first grade, functioning in English at a level equivalent to native English-speaking 3- and 4-year-olds. ESL students were performing at a level equivalent to native 3.6- to 4.6-year-olds.

Lum reported that, in the free-response situation, students tended to reply in the language most used by the teacher. That is, the more English was used in the classroom, the more English the students used in their replies. No data or analysis on this point are provided, however.

Lum looked at differences between classes within each instructional method on each of the five measures. There were no significant differences among the three TBE classes on any of the five variables. There was one significant difference between the two ESL classes on one of the five variables. In general, then, there seems to be little problem with Lum's data due to either nonrandom assignment selection bias or teacher effects (except in 1 out of 10 comparisons). Nevertheless, it would have been desirable if the author had taken more extensive steps to introduce additional statistical controls for these effects.

Lum also obtained student self-reports of use of first language outside the school. English performance was negatively related to use of L1 outside the school and use of L1 outside the school was positively related to being in a transitional bilingual education class.

Strengths

The project was restricted to monolingual Chinese speakers. There was a comparison of two types of instruction. Since most of the students were randomly assigned, the project was almost a true experiment. Extensive measures were used.

Discussion

Lum's categories are somewhat misleading since the ESL classes used Chinese in their subject areas, although to a lesser degree than the TBE classes. The generalizability of Lum's findings is limited for several reasons: the absence of measures of literacy and writing, a small sample composed of one grade in one school district, the unknown reliability of the method, and the young age of the students. However, the following conclusions are suggested by his study:

- English performance is a function of exposure to English.
- English as a second language alone (with the limited use of L1) was superior to a program of bilingual education plus English as a second language.

Lum's study did not present any data on learning in nonlanguage subjects. Some proponents of transitional bilingual education argue that instruction in L1 is critical in this area. Therefore, Lum's findings apply only to learning to speak English.

Name of Study : Models of Bilingual Education: Comparisons of Effectiveness

Authors and Date : Moore and Parr (1978)

Location : Small, rural town in West Texas

Treatment Group : 130 Spanish-dominant students with limited English proficiency

Comparison/Control Group : 77 English-dominant students

Duration : 9 months

Ages : Grades K-2

Type of Program : Variations of transitional bilingual education

Description

Moore and Parr (1978) studied 130 language-minority and 77 English-dominant students in grades K through 2 in four schools in a Title VII project. Four types of programs were represented: maintenance bilingual education, transitional bilingual education, minimal bilingual education (not more than 20 minutes a day of formal Spanish instruction), and non-bilingual (all-English) classes.

The school district was located in a small, rural community in West Texas where Spanish is the home language of roughly one-third of the students. Thirty percent of the students were from low-income families and many families were highly mobile because of seasonal work.

Moore and Parr summarize their results:

> Non-bilingual classes scored significantly higher than bilingual classes on measures of reading and language achievement in English. Because students were not randomly assigned to treatment groups, these results should be viewed with caution. Covariance was used to attempt to correct for pretest differences, but covariance systematically underadjusts for initial differences between groups.

In addition, there were no differences among the groups on the math scale of the CTBS. Additional analysis by the authors found that sex and an unspecified rating of teacher competence had significant effects on some of the measures.

Strengths

An analysis of covariance using socioeconomic status, language dominance, teachers, teacher competence rating, aide competence, and school was employed. Thus, this study has better statistical control than many studies having nonrandom assignment.

Discussion

There are some problems with Moore and Parr's analysis. Sex differences and levels of teacher competency should have been controlled. It is surprising, however, that classes with teachers rated as less competent generally had better results.

Name of Study : Acadiana Bilingual Bicultural Education Program: Interim Evaluation Report 1978-79

Author and Date : McSpadden (1979)

Location : Lafayette, Louisiana

Treatment Group : 142 Title VII students (blacks and whites)

Comparison/Control Group : 54 students (blacks and whites)

Duration : 9 months

Ages : Grades K and 1

Type of Program : Transitional bilingual education

Description

McSpadden (1979) reported on the first year of a French-English Title VII program in Louisiana where the project included both public and parochial school participants and controls in kindergarten and first grade. About 37 percent of instruction was in French in language arts, math, and social sciences. A bilingual specialist, associate teachers, and aides taught the French portion of the curriculum while regular classroom teachers in the bilingual classes taught only in English. The regular teachers performed support activities during the French instruction.

The method the school used to select the participant and control groups is not given, but analysis of covariance was used to adjust pretest differences. On a locally developed French language test of French skills, math, and social sciences, program participants had significantly greater gains over the school year than did nonparticipants.

On a standardized achievement test in English, there was no difference in the performance of the two groups on any of the various language skill subscales or on the math portion.

Strengths

McSpadden employed analysis of covariance to adjust for pretest differences, and his sample size was fairly large.

Discussion

The program improved L1 performance with no detrimental effect on English performance. On the other hand, there were no gains in the students' performance in English as a result of daily instruction in L1 (i.e., no facilitation effect). Therefore, since the criterion used for the present paper is one of improved performance in English or other subject matter, we conclude that the McSpadden study found transitional bilingual education to be ineffective, subject to the following limitations:

- It could take more than 1 year for the facilitating effect to occur.

- There are no data presented to show that students in the program were language-limited. Instead of being a language-minority group struggling to learn English, they could basically be an English-speaking group learning a second language (French).

- The facilitation effect is generally discussed in terms of literacy, not oral proficiency measured in young children. First-grade students are just beginning to learn reading and writing skills. The facilitation effect would not yet be applicable.

- The generalizability of the results is limited since the study deals only with kindergarten and first grade.

Name of Study : Acadiana Bilingual Bicultural Education Program: Interim Evaluation Report 1979-80

Author and Date : McSpadden (1980)

Location : Lafayette, Louisiana

Treatment Group : 203 Title VII students (black and white)

Comparison/Control Group : 60 students (black and white)

Duration : 9 months

Ages : Grades K-2

Type of Program : Transitional bilingual education

Description

McSpadden (1980) continued the evaluation of the Lafayette bilingual program into the next year of operation. One additional grade was added to the program, which now included kindergarten through second grade. This report does not compare the treatment group and comparison group for progress in L1.

Using the same analysis as the year before (McSpadden, 1979), the study compared English progress of the comparison and project students and found the following:

- There were no differences in kindergarten;
- Project (transitional bilingual education) students in grade 1 had significantly lower total reading scores than the comparison students; and
- Project students in grade 2 showed significantly poorer performance on word knowledge and math subtests.

Strengths

The author used a large sample size, longitudinal data, adjustment for pretest differences, and analysis of covariance.

Discussion

Although the results are subject to the same limitations discussed previously concerning McSpadden's 1979 report, they strengthen the interpretation that the program is not effective in developing skills in English and math. To the extent that grade 2 participants have not been replaced by student turnover, the grade 2 students represent a longitudinal cohort that has received 25 to 30 percent of all its math instruction over 3 years in L1. Although these students performed the same as the comparison group during the first year, an additional year shows they are beginning to fall behind.

Name of Study : Do Bilingual Education Programs Inhibit English Language Achievement? A Report on an Illinois Experiment

Authors and Date : Balasubramonian, Seeley, and De Weffer (1973)

Location : Illinois (excluding Chicago)

Treatment Group : 213 Spanish-speaking children in a transitional bilingual education program

Comparison/Control Group : 104 Spanish-speaking children in an English as a second language program

Duration : 5 months

Ages : Grades K-3

Type of Program : Transitional bilingual education, English as a second language

Description

Balasubramonian et al. (1973) compared 213 students in a bilingual program in 15 schools in Illinois with 104 Spanish-speaking children in ESL-only classes. The students were in kindergarten through grade 3.

All 317 children received instruction in English language arts as a regular part of the curriculum and an additional 30 to 40 minutes daily in special English as a second language instruction. It appears that the bilingual program consisted of one-half day in the traditional curriculum and one-half day in the bilingual program. English as a second language was part of the bilingual component. Thus, the bilingual program children were exposed to approximately 25 percent less English during the school day than were children who were taught English as a second language in the traditional curriculum.

The authors note that the rate of attrition was the same for the two groups over the school year and present a lengthy discussion of the application to their analysis of Campbell and Stanley's (1963) threats to internal and external validity. No differences in English-language performance were found between the ESL and bilingual-plus-ESL groups, leading the authors to conclude that the bilingual program was a success since the students improved their L1 skills (although no evidence on this point is presented in the paper) at no cost to their English performance.

Strengths

The study used a large sample size from several schools comparing two types of instruction. An analysis of covariance was run to control for nonrandom assignment. In addition, the authors examined pretest differences and found a significant pretest difference occurred in one of the three grades studied. Recognizing the possibility that the analysis of covariance underadjusted pretest differences, the authors cross-validated the analysis with a partial correlation analysis and used verbal and nonverbal IQ scores as an additional covariate for the students in grades 2 and 3.

Discussion

Given our criterion of improved performance in English skills, we do not find evidence for the success of transitional bilingual education in this study. No data were presented on progress in nonlanguage subjects. The supposed facilitation effect from learning Spanish (L1) to learning English did not occur, since the comparison group did just as well in English. However, since students were tested over only 1 school year, it may be that not enough time was allowed for the facilitation effect to become manifest. Alternatively, it may be that the facilitation effect in bilingual education leads to better English performance than would be found in a submersion program, but that ESL works even better and has a greater facilitation effect than the bilingual program.

Finally, the authors presented no information on how the instructional programs varied across schools. Systematic differences here could alter the perception of program outcome.

Name of Study : Bilingual Education in San Juan County, Utah: A Cross-Cultural Emphasis

Author and Date : Cottrell (1971)

Location : San Juan County, Utah

Treatment Group : K: 38 Navajo and Anglo students; grade 1: 53 Navajo and Anglo students

Comparison/Control Group : K: 46 Navajo and Anglo students; grade 1: 55 Navajo and Anglo students

Duration : 9 months

Ages : 5-7 years old (grades K-1)

Type of Program : Transitional bilingual education

Description

Cottrell (1971) evaluated a bilingual program for Navajo students in Utah which also contained Anglo students. Data from an unspecified source were given to show the low level of English usage in the homes of the Navajo students. Navajo college students were hired to form bilingual instruction teams with certified (non-Navajo) teachers. Navajo language was used to teach subject matter and a Navajo history and cultural program. English was taught through an English as a second language approach.

The comparison group was formed from Navajo students in the neighboring school district. Cottrell notes that Navajo students in the comparison school district had more exposure to English-speaking children outside school hours and that the comparison school students had historically outperformed students in the project schools.

Project students and comparison students showed no differences in oral English skills or in MAT scores.

Strengths

The author used analysis of covariance to adjust for preexisting differences between the treatment and comparison groups due to nonrandom selection.

Discussion

Since students from the comparison schools were historically known to outperform students from the project schools, the program effect was probably underestimated by the analysis of covariance. Furthermore, Cottrell did not separate program effects on Anglos from those on Indians. These two groups were differentially distributed in the study:

	Navajo	Anglo
Treatment	76	15
Control	54	47

The table above shows a significant difference in the proportions of Anglos and Indians in the groups. However, Cottrell's analysis of covariance does not seem to take into consideration this important fact.

Name of Study : The Effects of an English-Spanish Primary-Grade Reading Program on Second and Third Grade Students

Author and Date : Huzar (1973)

Location : Perth Amboy, New Jersey

Treatment Group : 84 randomly assigned Puerto Rican students

Comparison/Control Group : 76 randomly assigned Puerto Rican students

Duration : 2 to 3 years

Ages : Grades 2 and 3

Type of Program : Transitional bilingual education

Description

"When the program was initiated in 1969, the subjects were randomly assigned to either bilingual or regular classes, which remained intact throughout the primary grades" (Huzar, 1973, p. 34). The Inter-America Test was administered in 1972 to 84 program students then in grades 2 and 3"(two classrooms each), and to 76 control students (two classes each of grades 2 and 3). There were no initial differences between the two groups on the Metropolitan Reading Readiness Test or on an IQ test administered to the third grade. A posttest-only control group design (see Campbell and Stanley, 1973) was used, and the difference between the groups was tested by the t-test. "The posttest-only control group design is perhaps the only setting for which the t-test is optimal" (Huzar, 1973, p. 40).

When the two bilingual classes' scores were averaged there were no differences between the groups on their English reading skills. Within the bilingual group in one grade, the class having two bilingual teachers (one for Spanish, one for English) performed significantly better in English reading skills than the class having one bilingual and one monolingual teacher, but in the other grade there was no difference.

A one-way analysis of variance comparing scores by sex for treatment and control classes was significant. Although the control boys had the lowest average score of the four groups (treatment x sex), the Sheffee test for post hoc contrasts was not significant. Thinking the Sheffee test was too conservative, the author performed a t-test on the two groups of boys and found the treatment boys scored significantly higher than the control boys.

	Number of Boys	
Grade	Treatment	Control
2	18	25
3	23	18

Strengths

The methodology made use of longitudinal treatment and random assignment, with tests for initial equivalence between the groups. Teacher effects were examined to a degree.

Discussion

The implication that transitional bilingual education was differentially effective for boys and girls is probably not correct. Both grades were combined in the analysis and an inspection of the sample frame shows the following:

There were more older boys in the treatment group than in the control group. Since older students would be expected to score higher, this distribution across grades puts a disproportionately higher number of higher scoring third-graders in the treatment group. Therefore, the result shown in the t-test comparison could indicate nothing more than the unequal distribution of boys across grades within the two groups.

The limitations on generalizing from the study are that it covered a limited number of grades (two) in one school district, and a limited number of subject content areas (one) in only one language group.

Name of Project	: The Follow Through Planned Variation Experiment, Volume 4-A
Author and Date	: Stebbins et al. (1977)
Location	: 5 sites
Treatment Group	: 492 follow through participants (64% not white or black ethnicity)
Comparison/Control Group	: 568 regular English curriculum students (18% not white or black ethnicity)
Duration	: 4 years
Ages	: Grades K-3
Type of Program	: Transitional bilingual education

Description

The Follow Through (FT) experiment covered 13 different instructional models designed to link and follow through on special preschool programs into the early elementary years. One of the 13 models--SEDL--was a balanced program of bilingual instruction. Approximately three-quarters of the children were Spanish-speaking. Dual language skills were stressed in most curriculum areas. In the bilingual classroom model, the English-speaking children learned Spanish and the Spanish-speakers learned English. The Follow Through evaluation covered 3 cohorts over 4 years (K-3). However, the analysis in volume IV of the report is the analytical method finally settled on for the study and supersedes earlier reports.

The analysis is very complex. Fifteen outcome measures were used and each was tested with 10 different statistical models at each site. The 10 models were various methods of adjusting for the effects of nonrandom assignment. Background covariates included WRAT pretest score, first language, income, occupation, ethnic group, sex, age, between-site characteristics, type of preschool, mother's education, and MAT score.

The basic analysis strategy was to seek consistency among the statistical models. Although each model introduced unique artifacts into the result, consistent findings across models are probably robust and can be taken as evidence of true effects. The authors concluded:

> The performance of FT children in this model on the various measures of the outcome battery varies considerably among sites. Cross cohort comparisons also vary among sites. In general, FT children in this model perform as well on the cognitive conceptual skills tests as they do on the effective measures; overall, 12 percent of these measures have null effects. FT children perform somewhat differently in the basic skills domain, where 20 percent of the effects are positive and 57 percent are null. Positive or null effects tend to be concentrated within particular sites. This implies that the SEDL program has a

> wide range of effects by site, ranging from 100 percent positive (in Tulare in the basic skills domain) to 50 percent negative (in Philadelphia, also in the basic skills domain). Bock and Stebbins (1977)

The SEDL program is one of but three American projects we found that followed the sequence of first teaching literacy in L1 before teaching L2 literacy (SEDL, 1979). The program consists of a coordinated K-3 curriculum in keeping with the overall approach of Follow Through.

Strengths

The SEDL program had a fairly large sample distributed across several sites, cohorts for replication, and an extensive battery of outcome measures. A very complex statistical analysis was used, making this one of the more powerful of the nonexperimental studies.

Discussion

The results of the Abt (Stebbins et al., 1977) analysis are summarized in table 2-2.

TABLE 2-4. SUMMARY OF FOLLOW THROUGH RESULTS - BILINGUAL PROGRAMS

	Results (%)		
Site	Positive	Neutral	Negative
Philadelphia	0.67	88.3	11
Los Angeles	0	94	6
Tulare	51	49	0
St. Martin	0	89	11
San Diego (Texas)	7	79	14
Total	11.9	79.6	8.5
All Follow Through	12.8	67.6	19.6

Source: Stebbins (1977), p. 244, table A-4-2.

The table shows the proportion of positive, negative, and neutral findings from 150 analyses per site (15 outcome measures by 10 statistical analytic models). The 13 Follow-Through models in general had little impact in comparison to the regular school program. The SEDL program was, for the most part, about as effective as regular schooling with 80 percent of the tests showing no significant difference. Within the 5 SEDL sites, one site (Tulare) stands out for its positive effects. It would not be unreasonable to conclude that something happened at this one site that was effective. However, it is by no means clear that the effect can be attributed to bilingual education since the TBE program was also replicated at four other sites where the proportion of negative results is far greater than the proportion of positive results. We could find no indication in Stebbins et al. (1977), SEDL (1979), and Bock and Stebbins (1977) of why

Tulare was so different from the other sites, both within the SEDL model and across all Follow Through models.

In addition to the unknown site-specific confounding, there are other problems of confounded treatments in the data. As a Follow Through Program, SEDL also provided--

- Medical and dental services,
- Nutritional programs,
- Social services,
- Guidance and psychological services,
- Individual and small group instruction, and
- Coordinated K-3 curriculum.

The scope of the SEDL program can be appreciated by the program costs which were $800 per pupil or 62 percent over and above the average pupil cost. One could reasonably expect to get improved performance through any number of instructional methods if one was given a budget increase of 62 percent. Given the budget available to SEDL (and to the other Follow Through Programs), the question can be raised that since they had so much to work with, why did so little result?

Our interpretation of the data is that no reasonably convincing evidence for the effectiveness of bilingual education is to be found in the data.

Name of Study : A Sociolinguisitic Approach to Bilingual Education

Author and Date : A. Cohen (1975)

Location : Redwood City, California

Treatment Group : 45 Mexican-American children in 3 cohorts

Comparison Group : 45 Mexican-American children in 3 cohorts

Duration : 2 year study of a 3-year program

Ages : Grades K-3

Type of Program : Transitional bilingual education

Description

The theory often put forth justifying TBE stresses a sequence of initial literacy in L1 followed by the development of literacy in the second language. The Redwood City project is one of only three American studies where it is clear the prescribed sequence was followed. Two program students and 18 comparison students were retained during the years of the study. Males were twice as frequent in the program as in the comparison group. 81 percent of program participants had parents who were born in Mexico and 77 percent of the comparison group parents were born in Mexico. Parental permission was required of participants. An extensive battery of tests were administered to measure language proficiency, language use, math, academic aptitude, language attitudes, socioeconomic level, educational environment of the home, and demographic factors.

Almost half the comparison students were receiving special assistance through Title I (24 percent), ESL (18 percent), or tutoring (4 percent).

Across the three cohorts and the multiple-test battery, 100 F-tests on English proficiency were conducted. Of these 100 tests, 14 were significant with 11 showing superior performance by the comparison group and 3 favoring the bilingual program.

Based on ratings by parents and observers and on student reports, Cohen reports "The bilingual project did promote greater use of Spanish..." (Cohen, 1975, p. 226).

One of the three cohorts showed superiority for the program participants in math and in gains in nonverbal IQ.

Cohen summarizes the findings as "Mexican-American children who were taught the academic curriculum in Spanish and English for several years appeared to be as proficient in most English language skills as comparable Mexican-American children taught only in English" (Cohen, 1975, p. 163). "The Mexican-American children following the bilingual program performed as well as, or better than, comparison children on tests in a nonlanguage subject matter, namely mathematics" (Cohen, 1975, p. 236). "The Bilingual Project also had no apparent detrimental effect upon the academic aptitude of the Mexican-American children involved. In fact, that program seemed to have enhanced academic aptitude in the case of the youngest group" (Cohen, 1975, p. 237).

Strengths

The program is replicated through three cohorts. Analysis of covariance was used to adjust for the effects of nonrandom selection, following tests for parallelism of the regressions and high within group correlations between pre- and post scores. Longitudinal data and a large battery of tests covering IQ, English, Spanish, and math developments were administered.

Discussion

The author interprets the results by emphasizing the point that language minority children can develop their home language in school without worrying about ill effects on English performance. However, Cohen began with the argument that initial literacy in L1 would lead to better levels of L2 skills. This hypothesis, which underlies the rationale for bilingual education, was not supported by the data. Indeed, if anything the data show some negative effect on English development.

From our perspective, we find the programs' effect on English development to have been neutral to a little negative with mixed results in arithmetic.

Name of Study : Final Report of the Compton Unified School Districts Title VII Bilingual-Bicultural Project, September 1969 Through June 1975

Author and Date : C. Stern (1975)

Location : Compton, California

Study Populations : 213 students in a Title VII program

Duration : 1 year

Ages : Grades 4 through 6

Type of Program : Transitional bilingual education

Description

This is the final report on a 6-year Title VII project. The results of the first 5-year evaluations are summarized in insufficient detail to permit an assessment. The findings of the sixth evaluation year are presented in considerable detail.

The Title VII project operated in one school in two of three classrooms at grades 4, 5, and 6. A variety of information is presented and the report is noteworthy for the detail of program process information provided. Among the data presented are norm-referenced analyses, analyses of gains for participants with no controls provided, and grade-equivalent scores. As discussed in chapter 1, we can make little use of this data. However, the author also presents an analysis of covariance of the California Test of Basic Skills and its subtests for grades 4, 5, and 6. When averaged across all three grades, the program effect was negative on all 9 component scales of the CTBS. Looking at results across grades, there were 4 negative findings, and 5 that were no different in grade 4; 5 negative results and 4 that were no different in grade 5; and 8 negative results and 1 that was no different at grade 6.

Additional data are presented showing that students who have been in the program 3 or more years score higher than students who have been in 2 or fewer years.

Strengths

Stern employs some longitudinal analysis and adjustment for the effects of nonrandom selection by analysis of covariance. The program she studied had been operating long enough to become stabilized and involved a large number of participants.

Discussion

Generalization is limited by the fact that the study took place in only one school and in three grades. To a degree, the results are inconsistent; if longer exposure to the program leads to better performance, why did the controls do better? There are two interpretations for this situation. First, the analysis of covariance adjustment may not have been able to overcome the severe selection bias that occurs when two-thirds of the students are placed into the program. Therefore, the longitudinal analysis is correct. Alternatively, so much time is taken away from practicing English when students first enter the program that scores are depressed. As time passes and more English is added to the curriculum (as is characteristic of the TBE program), progress is noted. Nothing in the data enables us to select between these two alternatives.

Finally, Stern notes that after 6 years of program operations, some former participants were included in the comparison group. This situation creates a bias against the program if the former participants had been "graduated" from the program on the basis of improved performance. Unfortunately, Stern does not examine the possibility of such an effect's having occurred.

Name of Study : ESEA Title VII Bilingual Program

Author and Date : Carsrud and Curtis (1980)

Location : Austin, Texas

Treatment Group : Grade 4: 80 Title VII program students;
Grade 5: 92 Title VII program students

Comparison/Control Group : All other Mexican-American students in school district

Duration : 5 Years

Ages : Grades 4 and 5

Type of Program : Transitional bilingual education

Description

Carsrud and Curtis (1980) studied the change in percentile scores from the first grade for students in a Spanish-English bilingual program. They compared students who had been in the Austin, Texas, Title VII project for 4 to 5 years with students never in the project and with all Hispanic students in the district (see table 2-3). A significant proportion of the students (over one-half of the project fourth graders and one-third of the fifth graders) were monolingual in English. Many of the other students were English-dominant students. Thus, the majority of students served by the program were either English dominant or English-monolingual.

Comparison of gains for project and nonproject students were carried out by regression analysis with project status and pretest score as predictor variables. A significant difference in favor of the project was found in math for students who had been in the project for 5 consecutive years, but no differences were found for English or for both English and math in students who had been in the project continuously from grade 1.

Strengths

Carsrud and Curtis carried out a longitudinal analysis with pretest adjustment to correct for effects of nonrandom assignment.

Discussion

Although the longitudinal nature of the data is a plus, there is no doubt that there was attrition over the 4 to 5 years covered by the study. It would have been very desirable if the authors had provided an analysis of the effects of attrition.

It is worth reproducing Carsrud and Curtis' data table (see table 2-3) to illustrate the magnitude of the problem encountered in teaching many language-minority students. Most Mexican-American students began school at the national norm level. Note that in 4 or 5 years the Mexican-American students' performance has declined about 15 percentile points whether or not they were in the bilingual program.

TABLE 2-5. PERCENTILE SCORES OF EXPERIMENTAL AND CONTROL GROUPS

1980 Grade	California Achievement Test Scale		First Grade Percentile	1980 Percentile	Change
4	Reading	Project	54	50	-4
4	Reading	Nonproject	54	46	-8
4	Reading	District (Mexican-American)	54	39	-15
4	Math	Project	51	39	-12
4	Math	Nonproject	50	31	-29
4	Math	District (Mexican-American)	52	38	-14
5	Reading	Project	63	39	-24
5	Reading	Nonproject	59	32	-27
5	Reading	District (Mexican-American)	57	43	-14
5	Math	Project	67	38	-29
5	Math	Nonproject	55	33	-22
5	Math	District (Mexican-American)	58	39	-19

Name of Study : Evaluation of the Impact of ESEA Title VII Spanish/English Bilingual Education Program

Authors and Date : Danoff et al. (1977, 1978)

Location : National sample

Treatment Group : 5,800 Title VII Spanish-English project students

Comparison/Control Group : 2,400 non-Title VII students

Duration : 6 months

Ages : Grades 2-6

Type of Program : Transitional bilingual education

Description

The American Institutes for Research (AIR) carried out a national evaluation of the impact of ESEA Title VII Spanish-English bilingual programs (Danoff et al., 1977, 1978). The first results were released in February 1977 and caused a stir in the educational community because the research found little positive effects for students participating in bilingual education programs.

The AIR sample consisted of 5,800 project students in 37 sites and 2,400 non-Title VII students controlled for ethnicity, socioeconomic status, and grade level. The programs were in their fourth or fifth year of operation.

Results of the study indicated that Title VII students did slightly worse than the non-Title VII students in English language arts. In math, the Title VII students surpassed the control group. In response to widespread criticism of the study from the bilingual community and the National Institute of Education, additional data were collected on a subsample of the original universe. The subsample was posttested after a longer treatment interval than the original sample. The additional data were extensively reanalyzed by AIR (Danoff et al., 1978). This additional analysis somewhat changed the original conclusions: the reanalysis found no difference in math scores between the program and comparison students and continued to find superior English performance for the comparison group. Compared with national norms, both groups were in the bottom fifth of the nation in English and the bottom third in math.

Strengths

The authors drew a very large sample and had a well-controlled data collection program. The study included extensive sophisticated statistical analyses including analysis of covariance to adjust for preexisting differences due to nonrandom assignment and a statistical test to determine whether there was an underadjustment problem with the analysis of covariance.

Discussion

The AIR study has been criticized by proponents of bilingual education (Gray, 1977; Cardenas, 1977; O'Malley, 1978). Based on these criticisms, the National Advisory Council on Bilingual Education concluded, "The methodology used in conducting the study has been critically analyzed by various independent educational research and evaluation experts, who have rendered the findings of the study completely invalid." The key elements of these criticisms and our responses to them are presented below:

1. Comparison Across Programs. Bilingual education is very complex. The AIR study failed to acknowledge that bilingual programs develop and exist under varied conditions due to specific district characteristics: linguistic needs, availability of qualified teachers, adequacy of curriculums, district commitment, and political underpinnings (Gray, 1977). The AIR study failed to recognize these differences and treats bilingual education as an undifferentiated and uniform program. As a result, positive findings are cancelled out by any negative findings so that the effects of a good bilingual program are lost.

The criticism that it is unfair to compare across programs does not acknowledge the needs of policymakers to make informed decisions based on representative data. If transitional bilingual education is generally effective, its effects will show up on the average. The AIR study reflects both good and bad programs, many of which suffer from very real implementation problems and resource constraints. However, these contraints and conditions reflect actual district and school problems which determine how effective a program can be. Such programs must be included to measure the effectiveness of bilingual education.

2. Testing Interval. Only 6 of the 37 projects involved in the study were tested over more than a 6-month period. The remainder of the projects were analyzed for program results over shorter periods. Evidence of cumulative gains in bilingual education over several years indicates that such short periods would not allow observation of the real long-term improvements due to transitional bilingual education.

Danoff et al. (1978) reanalyzed the data taking a subsample of the original universe:

> The results can, it turns out, be summarized succinctly. For both the grade 2 cohort and the grade 3 cohort, the fall-to-fall achievement gains in English Reading and in Mathematics Computation in Title VII projects were neither significantly nor substantially different from what would have been expected without Title VII treatment, with one possible exception: the grade 3 cohort of Title VII

> children in communities for which no comparison classrooms were available (i.e., in the urban Northeast region) showed substantial gains in Mathematics Computation skills. In light of these trends across methods, the significant differences favoring non-Title VII in some of the grade 2 analyses of covariance were more than likely due to that method's tendency to undercorrect preexisting group differences in some situations.

Although the subsample is not totally representative of the original sample (excluding Southern California), the AIR study still is the most comprehensive study undertaken of bilingual education.

A number of studies in the literature that we looked at reported gains over a 9-month period. Over a 6 month period gains should also be made. It is unlikely that a sample as large as that of the AIR study would have failed to detect short-run gains if they were made.

3. Teacher Qualifications. The AIR Teacher Bilinguality Scale indicated that only half of the teachers involved in Title VII projects were proficient in English and Spanish. Only 26 percent of the teachers participating in the study had bilingual teaching credentials.

There is a paucity of qualified bilingual education personnel in the United States. Therefore, that AIR found only 26 percent of Title VII teachers in their sample to be qualified bilingual educators is not surprising. This finding reflects the actual situation in the United States (Reisner, 1981).

Moreover, the AIR analysis measured the impact of teacher characteristics on student performance. The study indicated that formal credentials were not related to performance.

4. Improper Comparison Group. AIR asked the principals of the Title VII schools to identify nearby schools with similar student bodies without Title VII programs to form the comparison sample. A number of schools were unable to identify comparison schools. Furthermore, there were initial differences in the level of language skills between the Title VII group and the comparison group.

This last criticism does not recognize AIR's use of analysis of covariance to adjust for preexisting differences between the two groups. If we conclude that, despite their efforts, AIR failed to produce an acceptable comparison group, then we must also reject all the studies that employed inferior methods for identifying comparison groups.

Rossi (1979) and NIE (1979) have analyzed the entire AIR project. While recognizing that the AIR study is not without problems, they must be kept in perspective. Rossi acknowledged that, given the problems inherent with evaluations when assignment is nonrandom, AIR did an adequate job. That the study may not have evaluated poorly implemented projects does not detract from its conclusions as to the effectiveness of Title VII programs. It may or may not limit drawing conclusions about transitional bilingual education. The AIR study is one of the best existing nonexperimental studies of bilingual education and its conclusions must be given some weight. We think the following conclusions are supported by the AIR study:

- Title VII programs have not been shown to improve students' performance in school. Math scores seem unaffected but English performance is worse.

- The largest, most comprehensive study of a bilingual education program ever undertaken found no evidence that the program is an effective way to meet the needs of language-minority children.

STRUCTURED IMMERSION

Name of Study : Bilingual Education of Children: The St. Lambert Experiment

Authors and Date : Lambert and Tucker (1972)

Location : St. Lambert, Quebec, Canada

Treatment Group : 64 monolingual English-speaking children from St. Lambert, Quebec; cohort I: 26, cohort II: 38

Comparison/Control Group : English Control I Group: 48 monolingual English-speaking children in the same school, in a regular English curriculum; cohort I: 22, cohort II: 26

English Control II Group: 54 monolingual English-speaking children from Montreal, in a regular English curriculum; cohort I: 26, cohort II: 28

French Control Group: 47 children from French-Catholic school in St. Lambert, in a French curriculum; cohort I: 22, cohort II: 25

Duration : 4 years (longitudinal study)

Ages : Grades 1-4

Type of Program : Immersion

Description

The prototype immersion project is the St. Lambert program in French for English-speaking students in Canada (Lambert and Tucker, 1972). In the St. Lambert project, middle-class, monolingual speakers of English were introduced to French in an immersion kindergarten program taught by bilingual, nativeFrench-speaking teachers; students continued to receive monolingual French instruction through the first grade. In the second grade, an English language arts course was introduced for 1 hour a day. Ultimately, 40 percent of class instruction was in English, including English language arts, art, physical education, and music taught by a native English speaker. Sixty percent of class instruction was in French and taught by a native French teacher. Although instruction was delivered in L2 (French), students could speak to and ask questions of the teacher in L1 (English). The teacher, however, always answered in L2.

There were two types of comparison groups. Two regular English classes were established as controls. They received instruction in English, except for 1 hour of instruction in French as a second language (FSL) each day.* A French curriculum class also was chosen as a control. This class received the regular French curriculum for French schools in Canada. The groups were carefully compared for equivalence using socioeconomic status, IQ, language achievement, and home background factors based on interviews in the home. (The children involved in the study came from middle- and upper-middle-class homes.)

The St. Lambert study followed two cohorts from kindergarten through grade 6. The results showed a general trend of L2 performance superior to that of the English control group and approaching that of the L2 monolinguals.

Math was taught in French, but tests were given in both French and English with no difference in scores resulting from the language of the test.

English (L1) performance was initially depressed during the first 2 years when there was no formal English instruction. However, when English instruction was introduced into part of the school day, L1 scores improved to normative levels.

At grade 4, the experimental immersion group was equal to the English controls in subject area achievement and intelligence, but slightly lower in English oral and listening skills. Their achievement scores in French were average compared with Montreal norms; their scores in oral skills were slightly lower than those of the French control group. The immersion group was retested in grade 6. The English control group and the experimental group had equivalent scores on the English exams administered, but the experimental group was not equal to the French controls on the French tests. The immersion students, although having made significant strides, did not have native fluency in the language. However, the experimental groups' L2 performance was far superior to that of the English controls receiving French as a second language instruction.

Strengths

Following the progress of two cohorts permitted replication of the results. Considerable background information was gathered on the students so that comparability of the groups could be tested, and covariates were used in the analysis to adjust for initial differences. An extensive battery of tests and comparison groups were used, providing considerable information on the students' development. The longitudinal design made it possible to determine long-term trends in development. Finally, the overall pattern of results is consistent with the authors' theoretical propositions and is difficult to explain by any more parsimonious alternative.

---

* French as a second language is the equivalent of English as a second language when L1 is English rather than another language. We therefore include FSL as an ESL program, since that makes the terminology consistent for the American case.

Discussion

We have some concerns about the methods used in the St. Lambert study. First, the sample size was very small, especially for a longitudinal study in which normal attrition (students moving in and out of a particular school) can be expected to reduce the sample size over the years. The first experimental cohort contained 26 students, reduced to 20 by the fourth grade; the second cohort began with 38 students, but dropped to 27 by grade 3. The voluntary nature of the program raises special concerns about attrition, since a plausible alternative hypothesis would be that students who were doing poorly in the program were pulled out by their parents, thereby biasing the program with students who, for whatever reason, did rather well in learning French. In addition to attrition in the treatment group, attrition in the French-speaking control group was so high by the second year of the study that the authors felt compelled to supplement the comparison with the less desirable norm-referenced design in French.

Program participants were all volunteered by their parents. The first chapter of this report discusses how such a process can introduce bias into a study. This type of bias could have been especially acute in the St. Lambert study, since the program was begun by the school only after intense parental pressure for a special program to teach Canada's official second language to their children. Recognizing the possibility of selection bias, Lambert and Tucker gathered considerable preprogram data on the students and their home environment and found there were differences among the various groups on several parental attitudes and on two of five indicators of socioeconomic status. The authors concluded these were primarily differences between the French- and English-speaking comparison groups, since the treatment group mean fell between the means of the other two groups in the analysis of covariance. However, to have been completely certain of this interpretation, the authors should have carried out post hoc contrasts of various pairs of means. It is important to note there were no differences in nonverbal IQ among the various groups of students.

English performance was depressed at the end of the first grade relative to the comparison group, but equal at the end of the second grade. The authors attributed the improvement to a facilitating effect on learning to read a second language (English, in this case) from having first learned to read another. This argument overlooks the fact that the students came from English-speaking homes where parents were concerned that both languages be learned. These children had 40 percent of their school day (including classes in art, music, and physical education) in English. Sixty minutes of formal English instruction per day in the second grade could have been enough to bring children of this background up to par without any facilitating effect.

The authors also tried to demonstrate language transferability. Following a study of bilingual college students who were better able to discriminate sounds in a third language than monolinguals were, the authors included a test of discrimination of Russian phonemes at the end of each grade. No difference was found between the program and comparison students.

Despite this direct rejection of the transferability hypothesis, the authors proposed a transferability explanation for several results. The authors also played down the depressed English performance of students after grade 1 and the occasional negative effects indicating mental confusion, which appeared in the yearly IQ tests.

In reviewing the IQ test results on the Lorge-Thorndike "not-belonging" subtest, Lambert and Tucker (1972) pointed out the mean for the experimental class was above the mean of the controls. Although, the authors argued that this measure shows the experimental children reaped benefits, the difference was not statistically significant and was only about one-half the magnitude of the statistically significant difference favoring the comparison students on the vocabulary subtests, a difference the authors dismissed as worthy of "no special attention" (Lambert and Tucker, 1972, p. 123).

The authors concluded there was "no native language retardation deficit of any sort" (Lambert and Tucker, 1972, p. 152), but two tests of English proficiency showed the comparison students to be significantly superior to those in the program (see Lambert and Tucker, 1972, pp. 147-48).

The Lambert and Tucker study has another important implication for the U.S. problem: learning is best accomplished in language-segregated settings. A little-noted substudy in the Lambert and Tucker study looked at the consequence of putting some native French speakers into the French immersion classrooms. It was expected these students would provide French-speaking role models for the students learning French and would improve their performance. However, just the opposite happened. French (L1) Performance of the immersion students declined. Apparently these students could not keep up with the native French speakers, who tended to monopolize the teachers' time. The implication is clear. Language-minority students, at least during the early stages of acquisition of English, should be separated from their English-speaking peers.

Despite some technical problems, the study is impressive. The degree to which the St. Lambert experience is generally applicable to the United States, however, is unknown. The two settings differ in the following ways (see Paulston, 1978; Tucker, 1980) and perhaps others:

- St. Lambert was a middle-class, suburban community. American bilinguals are almost exclusively of lower socioeconomic status in either rural or urban communities.
- L2 was a high-prestige language in the community. Most American communities place a high value only on English.
- L1 was the majority language of the culture. In the American setting, L1 is a minority language.
- The goal of the project was to develop bilingual students, persons fluent in both English and French. The emphasis of American policy is on English mastery.

Name of Study : Three Year Evaluation of a Large Scale Early Grade French Immersion Program: The Ottawa Study

Authors and Date : Barik and Swain (1975)

Treatment Group and Comparison/Control Group :

	Cohorts					
Grades	I		II		III	
	Exp.	Cont.	Exp.	Cont.	Exp.	Cont.
K	220	200	270	310	140	130
1	200	225	160	170	—	—
2	108	120	—	—	—	—

Duration : 3 years (longitudinal study)

Ages : Grades K-2

Type of Program : Immersion

Description

A second major evaluation of a French (L2) immersion program was carried out in Ottawa by Barik and Swain (1975). In this program only L2 was used for instruction in kindergarten and grade 1. In the second grade 1 hour a day of English (L1) instruction was added to the curriculum. The study spanned 3 years and three cohorts in several schools.

The study design was much like that used by Lambert and Tucker (1972). Students were tested in both languages and comparison groups received 15 to 40 minutes a day of L2 instruction in an ESL-type setting. The study found that kindergarten immersion students did better in L2 after 1 year of instruction than did ESL-type students after 2 years. L1 performance was depressed during the first 2 years, but recovered in the third year to normative levels. L2 performance of the immersion students in all grades and all cohorts was far superior to the comparison group and approached the national norm. (Since the students were around the 70th percentile in L1, median performance in L2 indicated they were still somewhat deficient in L2 given their ability level.) Unlike Lambert and Tucker (1972), Barik and Swain did not include a native-French-speaking comparison group but used test norms instead.

Strengths

Barik and Swain had a large study population about whom they gathered extensive data. Students were nonrandomly selected for the treatment groups and comparison groups, but preexisting differences were adjusted for by analysis of covariance. By following several cohorts of children, Barik and Swain were able to gather longitudinal data. Program results were replicated in all three cohorts examined and were consistent with the findings of Lambert and Tucker (1972), as well as immersion theory.

Discussion

Barik and Swain (1975) do not explain how students were selected to participate in the program, which raises a question of whether selection bias was present. However, their use of analysis of covariance is the standard procedure for taking into account the nonrandom selection process. A second reason why selection bias might not have been a problem is that the pattern of results is inconsistent with any of the alternative explanations based on selection bias. Selection bias holds that program students are initially either superior or inferior to the comparison group and that this difference continues through the evaluation. Neither of these patterns is found in Barik and Swain's (1975) or in Lambert and Tucker's (1972) immersion studies.

Since Barik and Swain's study design was similar to that of Lambert and Tucker, most of our general comments about Lambert and Tucker also apply here with an important exception. By drawing on an extensive experimental program through two large school districts and by selecting the comparison group from nonexperimental classes throughout the districts, Barik and Swain had a much larger sample available. Unfortunately, the authors present no analysis of attrition effects, which occur in longitudinal studies to some degree.

Name of Study : English-French Bilingual Education: The Elgin Study Through Grade Five

Authors and Date : Barik, Swain, and Nwanunobi (1977)

Location : Elgin County, Ontario, Canada

Treatment Group : 73 monolingual English-speaking students

Comparison/Control Group : 79 monolingual English-speaking students in a regular English curriculum; other control groups, including full French immersion classes

Duration : 5 years (longitudinal data)

Ages : Grades 2-5

Type of Program : Immersion

Description

Barik et al. (1977) report on the fifth year of a longitudinal study of a partial immersion program. While the immersion programs of Lambert and Tucker (1972) and Barik and Swain (1975) began with 100 percent of the school day in French (L2) for grades K and 1, then gradually shifted to a 60 to 40 percent division between L2 and L1 (English) at the upper elementary grades with all content subjects taught in L2, this partial immersion program used a 50-50 language split of French and English beginning in first grade. One language was used in the morning, the other in the afternoon.

Mathematics, music, and French language arts were taught in French by a bilingual anglophone teacher. Science, beginning in third grade, was also taught in French. English language arts, physical education, and other subjects were taught in English. The curriculum content was the same as that followed in a regular English program. Beginning in grade 3 to 5 each language component was taught by a different teacher with native or native-like command of French.

The treatment group seems to consist of four cohorts of one class each (n = 73). Several types of comparison groups were used. To assess L1 performance, each cohort was paired with one class of students in the regular monolingual English program from a similar school in the same school district (n = 79). L2 development was assessed by comparison with students in another school district (apparently the data reported by Barik and Swain, 1975) that had both a full immersion program and regular L1 instruction with one period a day of formal instruction in L2 language arts.

Analysis of covariance used nonverbal IQ and age as the covariates. Each year's data were analyzed separately, although the nature of the data suggests that a repeated measures design may have been more appropriate (see also Lambert and Tucker, 1972; Barik and Swain, 1975).

Strengths

In the Elgin study, Barik et al. had a large sample size about whom extensive data were gathered. Several different comparison groups were used to measure different achievement levels. Students were randomly selected for the comparison groups, and for preexisting differences were adjusted for by analysis of covariance. By following several cohorts of children, Barik et al. were able to gather longitudinal data by replicating the treatment.

Discussion

The results are complex and are most easily understood by reproducing the authors' summary tables. Apparently, the data on L2 performance (table 2 in the report) were not adjusted by the covariates and statistical tests are not reported. Therefore, we cannot accept the results of the L2 data and will limit our discussion to math performance.

The authors note that understanding the math results requires an understanding of the English (L1) results. Apparently in response to initially poor performance in L1 (relative to the students instructed monolingually in L1), the school moved to strengthen L1 instruction by reducing the time given math instruction (math instruction was all in L2). Further, the students received less total exposure to L2 than did the total immersion students. Therefore, it may have been more difficult for them to maintain normal progress in L2 math. Nevertheless, while the comparison students performed significantly better in math in 8 of 88 comparisons, 80 of the 88 comparisons show that the students who were taught math in L2 were not falling behind in math performance by comparison to those taught in L1. Since the treatment group did not fall behind in math skills on 90 percent of the comparisons made, it appears that instruction in L2 did not impede subject matter acquisition.

Assessing the study's design is difficult. On the one hand, one must admire the authors for the ingenious way they patched together a rather comprehensive design from various sources. On the other hand, exactly how this patchwork approach may affect the results due to preexisting differences is unknown. Neither is it clear that using age and nonverbal IQ as covariates can fully control for any (unknown) preexisting differences. The inconsistent results within grade across cohorts and within cohorts raise further questions. The authors may or may not be correct in their speculation that these patterns reflect changes in school policy. Since this is a major issue, it would have been better if the authors had presented some hard data showing that these policy changes had indeed taken place rather than merely speculating on the point.

TABLE 2-6. TREATMENT AND COMPARISONS SCORES IN L2, ELGIN FRENCH IMMERSION PROJECT

Performance in French, Grades 2-5

	Elgin PFI*				Toronto TFI[a]**	Ottawa TFI/RE[b]***		
	Gr. 2	Gr. 3	Gr. 4	Gr. 5	Gr. 1	Gr. 2	Gr. 3	Gr. 4
French Comp. Test Level 1 (max.=45)	12.47	26.55	27.29	36.31	31.69	TFI: 38.39 RE: 10.58	40.03 14.20	- 12.15
Test de Rendement en Francais Gr. 1 Level (max.= 30)	10.94				19.10	---------	TFI	----------
Gr. 3 Level (max.= 30)		8.71	11.87	17.59			15.96	
Gr. 4 Level (max.= 40)			13.87					26.06
Test de Lecture Gr. 2 Level (Max.= 19)		10.50				12.83		
Gr. 3 Level (max.= 28)		11.44	15.73	21.50			19.73	
Gr. 4 Level (max.= 42)			16.79	22.38				28.82

[a] Data taken from Barik and Swain, 1976b (unit of analysis = individual).

[b] Data taken from Barik and Swain, 1977 (unit of analysis = class). Amount of French instruction in RE program = 20 to 40 minutes a day from kindergarten on (varies among grades).

* Partial French immersion
** Total French immersion
***Total French immersion/regular English curriculum

SOURCE: Barik et al., 1977, p. 465.

Name of Study : ABC's: McAllen's Immersion System

Authors and Date : Pena-Hughes and Solis (1980)

Location : McAllen, Texas

Treatment Group : 78 comparably limited Mexican-American children

Comparison/Control Group : 78 comparably limited Mexican-American children in transitional bilingual education classes

Duration : 9 months

Ages : Kindergarten

Type of Program : Immersion, transitional bilingual education

Description

An English immersion program for Mexican-American students began recently in McAllen, Texas, and has reported gains in test scores, self-concept, and discipline (Pena-Hughes et al., 1980). The project is different from the previously mentioned immersion programs in that the students are of low socioeconomic status, there is little parental involvement, it is not a voluntary program, the school is in a rural area rather than in a large urban area, the children are not monolingual but are comparably limited in both languages, and they are language-minority children, not of the majority culture.

The project was begun in kindergarten during the school year 1979-80. Teachers, aides, and students (n = 156) were randomly placed in four experimental classes (n = 78) and four TBE control classes (n = 78). Students were controlled for IQ and socioeconomic status. Students in the experimental classes were taught the same curriculum as the control students but their teachers spoke only English to them from 8:30 to 1:30. Students, however, could speak in either language. Physical education and cafeteria personnel spoke only English to the experimental children. Spanish language arts was taught from 1:30 to 2:30 in the afternoon.

Test results on the Language Assessment Skills (LAS) test (a State-approved language proficiency test) indicate that students in the experimental classes have made significant gains over the controls. Even though students were randomly assigned, the authors carried out tests to determine pretest equivalence and found the experimental group scored significantly higher on the Spanish pretest, with no difference on the English pretest. Therefore, pretest differences were adjusted for using analysis of covariance in order to assess program impact. In English proficiency the control group made a gain of 30.23 points from pre- to posttesting while the experimental group made a gain of 43 points.

The gains in Spanish proficiency of the experimental group were also significantly greater than those of the control group. Children in the experimental program gained 30 points while those in the control group gained 24. Both the experimental and control groups made significant gains in both Spanish and English over the school year. Teacher observations indicated that children in the immersion program could be identified by their improved use of English as compared with the students in the bilingual education program.

Strengths

The study was a true experiment in which both teachers and students were randomly assigned to the two treatment conditions. Although only in its first year, the design was longitudinal and the sample reasonably large.

Discussion

The TBE program used the approach generally called concurrent translation; in this approach the teacher immediately follows statements initially made in L1 with the L2 translations or vice versa. We argue later that this approach is counterproductive because students tune out the language they least understand and, in effect, receive only half a day of instruction. Therefore, any conclusion drawn from the McAllen study that immersion is superior to TBE must be made conditional: immersion was shown to be superior to one method of bilingual education that employs a very questionable pedagogical technique.

The pattern of results could also be due to the Hawthorne Effect, although this seems unlikely because both groups were in their first year of school. Also, both groups of students were in a special program.

The McAllen project was one of the six true experimental designs we have found in the literature and as such the reported success should be given greater weight than the results of studies that were less well designed. However, the generalizability of the study is limited because the experimental children had completed only kindergarten and the experiment is limited to only one school district.

The ongoing experiment in McAllen indicates that immersion can indeed succeed in the typical American bilingual setting, and that perhaps the Canadian experience does generalize more than had been previously thought. The difference between immersion projects and submersion cannot be overemphasized. An immersion program does not involve simply placing language-minority children into an English-speaking classroom to sink or swim. Immersion is a carefully structured program of L2 instruction which presumes no prior knowledge of L2. The teachers understand student queries in L1 even though they reply only in L2. Immersion programs may to meet the needs of language-minority children. They require much work on the part of schools, but there is a payoff for the effort.

CHAPTER 3

INAPPLICABLE STUDIES

The preceding chapter discussed those studies we found to be applicable to our question in terms of (1) the issues addressed by the studies and (2) their methodological soundness. This chapter discusses the studies we did not find applicable. The discussion is limited to the reasons why we did not consider these studies relevant to the issue. While many of the studies doubtless address some other question(s) adequately, we are not undertaking a general review of the literature across all possible questions. We are attempting to answer some particular policy questions, and so are concerned only with studies that provide suitable data for these questions.

The most common reasons for deciding that a study was not applicable for our purposes were the following, alone or in combination:

- The study was designed to answer locally relevant questions that did not address our question.
- No control or comparison group was included. The study reported gains only for the program participants.
- The norm-referenced model was used.
- No statistical or matching controls were employed where assignment to the program was nonrandom.
- A posttest-only design with nonrandom assignment was used.
- Results were reported only in terms of grade-equivalent scores.

The bulk of this chapter presents our reasons for rejecting a number of studies that have been widely cited as evidence to support transitional bilingual education. At the end of the chapter we describe an English as a second language study of a project widely cited as evidence supporting ESL and our reasons for rejecting the study as not addressing our questions. Our final discussion presents a summary chart of studies which did not meet our methodological criteria and the reasons for rejection.

Transitional Bilingual Education Studies

The Chiapas, Mexico, Study

Description

Modiano's (1968, 1973) comparison of the Spanish direct teaching and the Indian native language approaches in the Chiapas highlands of Mexico is probably the most frequently quoted study in the area of bilingual education.

Since Modiano's study has been well summarized by Engle (1975), we quote Engle at length:

> The Chiapas educational situation offers a natural experiment because they have developed three types of schools. Two use a direct teaching method entirely in Spanish; these are the State and Federal schools. The third is run by the Instituto Nacional Indigenista (INI), and introduces instruction in the vernacular. The children begin in a preparatory grade. During this year they are taught reading in the vernacular using a global method, and are given oral Spanish drills. In the second year of schooling, children are introduced to reading in Spanish, this time by a phonics method. In Modiano's study two factors were examined: the effects of the two methods of instruction on learning to read in Spanish, and the role of the teacher in the community.
>
> The sample was 1,601 children from 26 schools with 42 teachers in three tribal areas. Villages (generally with one school each) were matched as nearly as possible on variables such as distance from a road, amount of food, climate of the schoolroom, and resistance [to] or acceptance of schooling. Careful observations were made of actual classroom procedures. These observations, contained in Modiano (1973), highlight the importance of teacher training for rural teachers. In addition, her comments about poor attendance, low morale on the part of the teachers, lack of materials, and isolation of the villages, underline the need for analyzing the ecological situation of the school for potential alternative explanations.
>
> The assessment instrument, a Spanish language reading test, was developed specifically for this area; items were made relevant to the Chiapas Indian child. Reliability and validity estimates were adequate. The test was given to all children who the teachers considered knew Spanish well enough to take the test (about 30 percent). The children were not equated on age or IQ, since it was virtually impossible to obtain that information.
>
> Modiano (1968) compared the INI schools and the State and Federal schools on a number of variables. The INI school teachers identified significantly more children whom they thought could read Spanish with some understanding. Their judgments were corroborated by the finding that the INI students also scored significantly higher on the Spanish reading test.

Reasons for Rejection

There are several major reasons for rejecting Modiano's study. First, and most important, we have reason to believe the comparison group was not equivalent to the treatment group. The process of assigning students to either the Indian or State schools was not random. Variables known to affect language learning were not statistically controlled. Modiano acknowledges she was unable to control for age differences.

Another problem with not controlling for preexisting differences lies in Modiano's acknowledgement that there was higher adult literacy in the treatment villages. Modiano proposes that the literacy rate was the result of adults having participated in the native language program. However, because of the recent implementation of these programs this proposition is doubtful. A possible alternative explanation would be that these communities had higher literacy rates even before the advent of the Indian schools. These literacy rates would be expected to affect village children's literacy rates and should have been controlled. Modiano seems to have made no statistical adjustment for these preexisting differences.

A further problem in Modiano's study is with the method of L2 instruction utilized by the Federal schools, which Engle describes:

> The Direct Method Approach advocates fluent second language learning before reading is introduced. Modiano's (1973) descriptions of the State and Federal schools, which were using only Spanish, indicates that the children were far from fluent in Spanish before reading instruction was begun. Further, no systematic oral Spanish instruction was attempted. Children learned by rote means and were generally confused. Thus, her results suggesting that the native language approach in INI schools [was] superior are not surprising. The study does not present a comparison of a good use of the Direct Method with the Native Language Approach. The variable of the teacher—his background, ethnic identification, training, and relationship with the community—has been too infrequently studied. Modiano (1973) assessed the relative effects of the native teachers (promotores) and the State and Federal teachers, primarily from the dominant culture, on the community in which the school was located. Promotores are usually sixth grade graduates from an Indian community who receive training. She used three measures of the effect of teachers on the community: the number of teacher-sponsored projects (this is one of the roles the teacher is supposed to fulfill); the percent of girls in school; and the percent of females which the teachers indicated knew enough Spanish to be tested. Chi-square analyses on each variable between the two kinds of school were significant, favoring INI schools. Indians also unanimously preferred Indian teachers. Thus, it appears clear that either the Indian teachers, or the vernacular in the schools, has some effect on the rest of the community, and that this effect is greater than the effect of the mestizo (non-Indian) teacher on the community.
>
> Is the difference between the schools a function of the language method, or of a sensitivity to the culture the Indian teacher brings that a mestizo teacher would not have? Two Indians in her sample teach in the Direct Approach schools. The children learned more from the two Indians than from the mestizos, but less than was learned by children in INI schools. A study of this question is needed.

The Rock Point Navajo School Program

Description

The bilingual program at the Rock Point Navajo school has been widely reported (Rosier and Holm, 1980; Rosier, 1977; Rosier and Farella, 1976; Voiri and Rosier, 1978). Following the successful development in the early 1960's of a new curriculum in English for Navajo students at Rock Point, which was then adopted by other Navajo schools, Rock Point began experimenting in 1967 with a bilingual program. Initially funded by Title I, the bilingual program "was limited to beginner (pre-first grade) level. Children continued to receive concentrated oral English instruction using an ESL approach but initial literacy--reading readiness--was introduced in Navajo" (Rosier and Farella, 1976, p. 379, emphasis added). In 1971 a Title VII grant expanded the bilingual program to include grades K-6 and "allowed the school to develop a comprehensive bilingual curriculum. The program was expanded from just reading readiness in Navajo to complete initial literacy in Navajo. After the children learned to read well in Navajo, they were introduced to English reading (during grade 2 or 3)" (Rosier and Farella, 1976, p. 380). However, a lapse in bilingual education resulted for those children who had graduated from the kindergarten program and completed the first grade before 1971.

The evaluation is based on three different analyses. First, it compares students enrolled in grades 2 through 6 of the Rock Point bilingual program with two comparisons groups drawn from other Navajo schools. Rock Point students were tested with the SAT in 1970, 1975, 1976, and 1977 and with the MAT in 1976. One comparison group was tested in 1975 with the SAT and the other was tested in 1976 with the MAT.

Rosier and Holm conclude that Navajo students who received bilingual instruction did better on standardized achievement tests than did Navajo students at comparable schools who had received English-only instruction. They also conclude that these students did better than an earlier cohort of Rock Point students who had received instruction only in English.

Second, a comparison is made between Rock Point students in the bilingual program and Rock Point students before the program began and it is reported that the bilingual students score better than the earlier group.

Third, a comparison is made between one group of fourth-grade students who were in the bilingual program continuously from kindergarten and a group of fourth graders from the same year who had had their bilingual program sequence interrupted in grades 1 through 2. It was found that the students with continuous bilingual education performed better than the group whose bilingual instruction was interrupted.

Reasons for Rejection

One problem plagued all three analyses. Each analysis was conducted in grade-equivalent scores. We noted in our first section the unsuitability of grade-equivalent scores in program evaluation; the Rock Point evaluation is a good illustration of one of the problems. In 1976 the Rock Point students were tested with both the SAT and the MAT. Growth rates in grade-equivalent scores were calculated for both tests. If grade-equivalent scores work, they should show the same growth rate on both tests. Instead, their growth rate was 1.18 on the SAT and 0.84 on the MAT (Rosier, 1977).

The first analysis which compared the Rock Point bilingual program students with groups from other Navajo schools had serious methodological problems. Rosier and Holm tried to match the treatment and comparison groups but we have doubts the comparison schools were similar to the Rock Point student body. The adequacy of the comparison between Rock Point and other schools depends on making a good match between the treatment and comparison groups. Rosier (1977, p. 13) reports the comparison schools were selected from other Navajo schools using the "Direct Method" but notes, "There were only a few BIA schools with such programs." Further, more than one-third of the comparison schools received students at grades 3 and up from Federal schools whose instructional programs were not known.

More important, Rosier and Holm (1980) note that the Rock Point school has historically outperformed other Navajo schools. They showed that the control schools scored higher than other Indian schools and interpreted this finding to mean a positive outcome for participation in the bilingual program. To arrive at this conclusion, Rosier and Holm should have shown that Rock Point and the control Indian schools were historically equivalent. This demonstration is particularly critical in view of Rock Point's historical superiority (Willink, 1968). However, Rosier and Holm failed either to test for equivalence or to apply statistical adjustments for differences.

Table 3-1 further illustrates Rock Point's history of academic superiority over other reservation and BIA schools. In the table we extracted the Navajo Area Norm (1970 comparison group) and the performance of Rock Point students before the program was instituted from Rosier and Holm's chart 12. Note that the same pattern that the authors interpret as evidence of the effectiveness of the bilingual program in the 1975 data is found in 1970, before the program began. In both years, Rock Point students began below the comparison group at grade 2 and were above by grade 6.*

---

* The relative size of the sixth-grade differences between 1970 and 1975 could be evidence of a program effect. However, other possible reasons for the increase are:

- o Different tests were used in 1970 and 1975.
- o Different comparison groups were used.
- o The nature of grade-equivalent scores makes such a comparison of doubtful validity.
- o The size of the difference is less than the year-to-year differences sometimes found between consecutive classes at the same grade.

TABLE 3-1. ROCK POINT AND COMPARISON SCHOOL PERFORMANCE, IN GRADE-EQUIVALENT SCORES

	1970			1975-1977		
Grade	Rock Point	Comparison	Difference	Rock Point	Comparison	Difference
2	1.8	1.9	- .1	1.94	2.16	- .22
4	3.9	3.2	.7	2.67	3.13	.46
6	5.4	4.4	1.0	6.2	4.16	2.04

Finally, we note the data Rosier and Holm used to demonstrate their findings were incomplete. The 1975 Rock Point sixth grade was eliminated because there were too few students (six). The SAT comparison group was tested only once (1975), but the authors compared 3 years of the experimental testing program (1975-1977) to this one comparison test year.

In the second analysis, grade 6 students in the Rock Point bilingual program were tested with the 1973 SAT and were compared with earlier (1970) Rock Point students who had been tested with the 1964 edition of the SAT. The students in the bilingual program were found to be superior, but again there are difficulties with this analysis. It is not at all clear that the bilingual sixth graders did indeed receive initial reading instruction in Navajo. It is also not clear what effect taking different tests had on the scores or how successful published tables equating grade-equivalent scores across different editions of a test really are.

One the of most serious problems is the lack of analysis comparing similarity of cohorts between the program and 1970 cohorts. Serious problems can occur with comparing different cohorts of students within the same school and assuming they are equivalent when the number of students is small. This works only if you can assume each succeeding cohort of students is the same. Although the law of large numbers says this is the case for the entire population, everything else being equal, one or two schools are not the entire population, and chance fluctuations can produce considerable differences in ability and performance between any two successive classes. This point is illustrated in the Rock Point data, where two successive fifth grades (1976 and 1977) had average scores of 5.66 and 4.51, about a 25-percent difference in performance from 1 year to the next. An examination of test scores indicates a variation in the number of children being tested. This variation can account for cohort differences being found. The authors did nothing to control for cohort differences in their analysis.

In their final analysis, Rosier and Holm (1980) describe a comparison between two groups of fourth graders, showing the results of continuous versus interrupted bilingual instruction. The interrupted group began school with Navajo reading readiness instruction (apparently in kindergarten) but then entered the all-English first-grade program. In the third grade they were returned to the bilingual program. Rosier and Holm argue that these data provide an important evaluation of the program. They attribute the higher scores of the bilingual education students to their participation in the bilingual program. There is another interpretation to these findings, however. There should be little wonder that children who began schooling in one language, were changed to a second language, and then changed again to a mixture of both languages, all within 4 years, did not do so well as students who had a coordinated exposure to the two languages. Further, Rosier and Holm failed to examine equivalency between the two groups.

Finally, Rock Point is but one of three American studies to use the sequence of first teaching L1 literacy and then teaching L2 literacy. The facilitating effect on L2 that is hypothesized by first teaching L1 literacy is one of the major justifications for TBE. Consequently, Rock Point is a major study. However, the adequacy of Rock Point as a test of this hypothesis is doubtful since not all the classes in the bilingual program seem to have been first taught literacy in L1. One of the most puzzling aspects of the study focuses on how grades 4, 5, and 6 were taught initial literacy. Rosier and Holm (1980) state that the fifth and sixth graders were taught initial literacy in Navajo in grade. 1. But if the dates reported by Rosier and Farella (1976) are correct, there was no bilingual instruction other than kindergarten reading readiness prior to 1971. By the 1975 testing, students who had been first graders in 1971 would have been in grade 5, so that of the three grade 6 classes, only one would have been taught initial literacy in Navajo and of the three grade 5 classes, only two would have been taught in Navajo. Since it was the grade 5 and 6 classes who scored the largest gains over the comparison groups, it is not clear how the results should be interpreted.

Finnish Immigrants in Sweden Study

Description

A recent study of Finnish immigrant children in two Swedish school systems (Skutnabb-Kangas and Toukomaa, 1976) is fast becoming one of the most widely cited studies supporting the use of L1 in the schools (see Pifer, 1979; Troike, 1978; McConnell, 1980 a, 1980b; Baral, 1979; Rodriguez-Brown and Junfer, 1979; Cummins, 1980).

It is generally thought that the study by Skutnabb-Kangas and Toukomaa (1976) shows that the more schooling in Finnish (L1) children had before beginning instruction in Swedish, the better their Swedish (L2). It is then inferred that this finding supports the use of L1 in the United States for children from non-English-speaking backgrounds.

Reasons for Rejection

We rejected the Skutnabb-Kangas and Toukomaa (1976) report for several important reasons. The report lacks much of the detail needed to figure out exactly what the researchers did. In addition, the authors did not have random assignment, and they did not try to match the comparison and experimental groups or to control statistically for preexisting differences. Further, there is a high attrition rate in the authors' sample.

Missing data pose a particulary severe problem. The authors state that the study covered 687 students (Skutnabb-Kangas and Toukomaa, 1976, p. 48), but their key analysis is based on only 150 students. When almost 80 percent of the data are missing, serious distortions can be introduced into the study. The authors give no consideration to this problem.

Skutnabb-Kangas and Toukomaa presented no statistical analysis of many of their data. When we did the statistical analysis, it did not support the conclusions generally drawn from the study. This section closely examines the Skutnabb-Kangas and Toukomaa report and argues that the conclusions the researchers draw from their study are unwarranted.

Length of Residence in Sweden. Skutnabb-Kangas and Toukomaa (1976) concluded that the 11- and 12-year-old children who had had some prior schooling in Finland performed as well, relative to the norm, as did 7- to 8-year-old immigrant children living in Sweden from an early age. The implication attributed to this analysis is that initial schooling in L1 leads to better L2 performance. The authors conclude the "learning potential in the foreign language (Swedish) is influenced by ability factors, but also by their skills in the mother tongue; in other words, the better a pupil has preserved their mother tongue compared with others who have lived an equal length of time in the receiving country, the better are their prerequisites for learning the foreign language" (Skutnabb-Kangas and Toukomaa, 1976, p. 78).

We suggest that the data are open to alternative interpretations. Krashen (1979), in a recent review of the literature on age and L2 learning, concluded that older learners acquire the initial stages of L2 more rapidly than do younger learners, but younger learners ultimately reach higher performance levels. Krashen's findings lead to a very different conclusion about the educational program for language-minority students. While the implication that has been drawn from Skutnabb-Kangas and Toukomaa is to use L1 in the early years of schooling, the implication from Krashen is to maximize use of L2 in the early years. Language educators focusing on the pattern reported by Krashen have stressed the importance of beginning second-language instruction at the earliest possible age.

It would be necessary to control for both age and L1 proficiency before the implication that language minorities are best taught first in L1 could be demonstrated. Skutnabb-Kangas and Toukomaa did not do this.

L2 Test Performance and Schooling in L1. Perhaps the most important data in Skutnabb-Kangas and Toukomaa's report on those reproduced here in table 3-2. These data have been widely interpreted as showing that L2 performance is better the longer the exposure to L1: third to sixth graders with 3 or more years of school in Finland performed better in Swedish than did those with 1 to 2 years of school in Finland, who, in turn, did better than those schooled entirely in L2. (Points refers to three categories of performance in Swedish, with 1-2 being the low end of the scale.)

TABLE 3-2. RESULTS OF WRITTEN COMPREHENSION TESTS ACCORDING TO LOCATION OF SCHOOL (JAUHO AND LOIKKANEN 1974)

Percentage Points	Attended school only in Sweden		Attended school in Finland	
	Swedish-lang. class	Finnish-lang. class	1-2 yrs.	3 yrs. or more
1-2 (-)	12%	5%	14%	6%
3	26%	40%	11%	12%
4-5 (+)	62%	55%	75%	82%
Total	100%	100%	100%	100%
N	65	40	28	17

SOURCE: Skutnabb-Kangas and Toukomaa, 1976.

The authors present no statistical analysis of the data. We carried out a lengthy statistical analysis which is presented here to examine the extent to which these data support the use of L1 instruction with language-minority children. The analysis is complicated by two factors. Interpreters of Skutnabb-Kangas and Toukomaa have assumed two different underlying models of the language learning process. Some models assume a continuous underlying function—"the longer the Finnish children were educated in Finnish, the better their academic achievement was in courses taught in Swedish" (National Clearinghouse for Bilingual Education, in Skutnabb-Kangas, 1979). Other models assume a discontinuous, step function; Cummins' threshold hypothesis (1976) argues that a certain level of L1 skills must be attained before a facilitating effect of L1 on L2 will occur. The applicability of various statistical tests to the data depends in part on whether the underlying function is continuous or has a step. The step function implies a dichotomy in the data and underlying function and the use of categorical level tests. The continuous function implies that the three conditions—school in Sweden, 1 to 2 years in Finland, and 3+ years in Finland—form an ordered metric and tests suitable for ordinal data are appropriate.

The second complicating factor is that the small sample (n) in some cells of the table raises questions about the suitability of the χ^2 test. Consequently, our analysis will include χ^2, likelihood ratio χ^2, asymmetric λ, Somers's d, and gamma values. (The coefficient divided by the asymptotic standard error can be treated as an estimate of the t-distribution to test the significance of the coefficient.) Table 3-3 presents the statistics calculated.

TABLE 3-3. SUMMARY OF THE STATISTICAL ANALYSIS

Test	Analysis 1	2	3	4	5	6	7	8
χ^2	11.2	4.79	2.59	2.65	3.49	2.59	7.7	3.08
Probability	.08	.31	.27	.26	.17	.11	.02	.21
Likelihood ratio ratio^2	11.7	5.11	2.83	2.94	3.66	2.83	8.3	3.18
Probability	.07	.26	.24	.23	.16	.09	.02	.20
λ	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Somers's d	.09	.14	.21	.196	.15	.21	.140	-.03
Approximate probability less than	NSD	NSD	NSD	NSD	.05	.05	NSD	NA
Gamma	.186	.29	.46	.437	.36	.49	.276	-.05
Approximate probability less than	NSD	NSD	NSD	NSD	.05	NSD	NSD	NA

Note: NSD = No significant difference

Each column of table 3-3 presents the results of an analysis corresponding to one possible interpretation of the data. The probability levels for Somers's d and gamma are approximations to the t-distribution based on the asymptotic standard error. In our opinion, generally the most useful statistic shown is the coefficient, which gives the probability of knowing in which category of Swedish performance a student will be found, given the student's exposure to school in Finland.

Our analysis is as follows:

1. The first column in table 3-3 presents the analysis of all the data shown in table 3-2. In addition to the statistics presented in table 3-3, r, τ and Spearman rank-order correlation coefficients were computed. The probability value of all three fall within the range 0.10 ± 0.02. The results are quite clear. No matter what assumptions are made about the metric and the appropriate test, there is no relationship between L1 and L2 performance to be found in the overall data.

2. We analyzed the data excluding the group in the bilingual program in Sweden. The theoretical reason for dropping these students from the analysis is that this group introduces another variable (bilingual education) into the analysis. Also, since this is the worst scoring group, eliminating it maximizes the opportunity to confirm the conclusion. As can be seen in column (2) of table 3-3, the battery of nonparametric tests found no significant association in the data. As a further test, both the Kruskal-Wallis test and a one-way analysis of variance were performed. These statistics were insignificant ($P<0.4$ in both tests). To further explore the possibility of association between the two variables, the Pearson r, Spearman rank-order, and τ correlation coefficients were calculated. All were insignificant and in the range 0.15 ± 0.02.

This analysis is the best test of the underlying model that prior exposure to L1 instruction is a continuous function showing a facilitating effect of L1 on L2. There is no evidence in Skutnabb-Kangas and Toukomaa's data to support such a conclusion.

Further, this analysis is probably the best test of the threshold hypothesis since none of the proponents of that hypothesis seems to identify a school year where the threshold is found. When examining the full range of data, no glitch is found that would correspond to the presence of a threshold.

3. Inspection of the data suggests the threshold may have been passed by only the group with 3 or more years of schooling in Finland. To test this hypothesis, the group with 1 to 2 years of schooling in Finland was eliminated and the group with 3+ years was compared to the group schooled entirely in Sweden. Again, there is no evidence of a threshold or of a facilitating effect of L1 on L2.

4. To further explore the threshold hypothesis, students with 3+ years of school in Finland were compared with those with 1 to 2 years of school in Finland combined with those schooled only in Sweden. Once again, there is no evidence of a threshold effect.

5. In their discussion, Skutnabb-Kangas and Toukomaa (1976) refer to the "level of achievement of normal Swedish pupils (4 or 5)." Following this distinction, we combined the two lowest skill level categories (called points in table 3-2) and compared the three groups with skill-level dichotomized.

Only gamma and Somers's d showed small, but significant effects. However, the use of these ordinal metric statistics with a measure having only two categories is problematic. A more appropriate measure in this case is τ_c, which was also significant $\tau_c = 0.17$($t= 1.98$, $P<0.05$).

6. Another comparison specifically mentioned by Skutnabb-Kangas and Toukomaa (1976, p. 66) is produced by eliminating the group with 1 to 2 years of school in Finland from the data analyzed in No.4 above. In this comparison only Somers's d was significant. The more appropriate τ_B was not significant ($\tau_B = 0.17$, $P>0.1$).

7. Skutnabb-Kangas and Toukomaa (1976) note that "Two years in a Finnish class in Sweden did not... make for as good a basis for learning Swedish as the corresponding time in Finland." Although the χ^2 is significant, it is a questionable test in this case since over 20 percent of the cells have an expected value of less than 5. More important, none of the measures of strength of relationship—especially the asymmetric χ^2—indicates the presence of a relationship. Even if a significant relationship is correct, note the distribution: students schooled in Finland are more likely to score both higher and lower than those schooled in Sweden. Everything considered, we find no support for the authors' claim in the data.

8. Finally, since the Skutnabb-Kangas and Toukomaa study is cited in support of bilingual programs, we should look at what happened to the students in the bilingual program in comparison with the students in all-Swedish classes. Although the direction of the data is against the bilingual program, there is no significant difference in the performance of the two groups.

The major problem we have with the correlation analysis is that we cannot interpret table 3-3. The most it seems to show is that good students tend to perform well in any number of subjects, including languages. Skutnabb-Kangas and Toukomaa go on to discuss two aspects of the table.

We have taken Skutnabb-Kangas and Toukomaa's table apart in a number of ways in an effort to find something supporting their argument. Looking across the eight analyses, one is simply overwhelmed by the lack of relationship expressed. Only two analyses found possible significance (ASE only approximate the true estimate of the standard error, and borderline cases of significance should be treated as doubtful) in the strength of relationship if the variables can be assumed to be ordinal. The important question to ask about the authors analysis is, "Why one would create such combinations to begin with?" Skutnabb-Kangas and Toukomaa offer no explanation. If certain combinations of the data are of theoretical importance, they should have been specified in advance (which the authors did not do).

In short, we find overwhelming evidence in Skutnabb-Kangas and Toukomaa's data for no relationship between L2 performance and development of L1.

The Correlational Analysis. Table 3-4 reproduces the third of Skutnabb-Kangas and Toukomaa's major analyses. The authors note that the absence of a significant correlation between age and the level of Swedish-language skill among students who immigrated at ages up to 5 years old is the result of their development in L2 reaching a plateau. Other data presented in the report show that this plateau is a higher level of L2 than that attained by any other group. Readers are cautioned to keep in mind that the presence of a correlation in table 3-4 does not necessarily imply anything about ability in Swedish. Since the students who had lived longest in Sweden (the 0-5-year column) had reached a plateau that the other students were still approaching, the higher correlations for the other groups indicate that until the plateau resulting from spending a long time in Sweden

TABLE 3-4. CORRELATIONS WITH SWEDISH-LANGUAGE SKILLS IN THIRD TO SIXTH GRADERS (OLOFSTROM)

	Age on moving to Sweden			All partial corr. (time held constant)
	0-5 yr.	6-8 yr.	9-11 yr.	
Age	.19	.58^{xxx}	.41^{xx}	.26^{xxx}
Sex (m)	.03	.12	-.23	-.04
Length of residence	.19	.68^{xxx}	.31^{x}	-
Age on moving	-.10	-.27^{x}	.00	.23^{xx}
Picture vocabulary (Finnish)	.51^{xxx}	.15	.43^{xx}	.41^{xxx}
Synonyms (Finnish)	.22	-.13	.24	.20^{xx}
Word Groups (Finnish)	.39^{xxx}	.01	.43^{xx}	.32^{xxx}
General level of Finnish	.37^{xx}	.03	.35^{xx}	.33^{xxx}
Observation speed	.27^{x}	.02	.55^{xxx}	.29^{xxx}
Addition	.44^{xxx}	.49^{xxx}	.50^{xxx}	.44^{xxx}
Better language (Swedish)	.05	.69^{xxx}	-	.03
N	68	48	49	165

Note: xxx, xx, and x refer to significant coefficients.

SOURCE: Skutnabb-Kangas and Toukomaa, 1976.

is reached, older children learn L2 faster than younger children, holding age on moving constant.

In 1979, the National Clearinghouse for Bilingual Education published a paper by Skutnabb-Kangas which refers to and summarizes the more extensive presentation in Skutnabb-Kangas and Toukomaa (1976). Table 7 in that document is reproduced here as table 3-5 showing oral performance in L2 (note that table 3-2 refers to written skills).

The listening comprehension test data suggest that longer schooling in L1 is related to L2 performance. Table 3-4 duplicates the analysis described above for table 3-3 on the data presented in table 3-5. The results are generally significant. However, it is not clear that these data can be taken as supporting the conclusions generally drawn from Skutnabb-Kangas

TABLE 3-5. LISTENING COMPREHENSION (SWEDISH) AND THE LANGUAGE OF SCHOOL ENTRY

	Group 1 Started school in Sweden		Group 2 Started school in Finland School years in Finland before emigration	
Test Mark (%)	Swedish classes	Finnish classes	1-2	3 or more
1-2 (-)	12%	4%	14%	12%
3	50%	33%	17%	12%
4-5 (+)	38%	63%	69%	76%
TOTAL	100%	100%	100%	100%
N	82	49	29	17

SOURCE: Skutnabb-Kangas, 1979.

TABLE 3-6. SUMMARY OF THE STATISTICAL ANALYSIS

Test	Analysis							
	1	2	3	4	5	6	7	8
χ^2	20.3	16.5	6.4	4.96	13.8	.853	3.9	8.53
Probability	.002	.002	.04	.08	.001	.003	.14	.014
Likelihood ratio^2	22.1	17.8	7.4	5.77	14.2	8.7	3.9	8.77
Probability	.001	.0023	.02	.06	0	.003	.13	.013
λ	.12	.156	0	.017	.31	.36	0	.145
Somers's d	.223	.234	.25	.066	.31	.56	.019	.222
Approximate probability level	.01	.05	NSD	NSD	.01	.01	NS	.05
Gamma	.385	.435	.44	.397	.58	1.43	.04	NA
Approximate probability level	.01	.05	NSD	NSD	.01	.01	NSD	NA

and Toukomaa (1976), since that report dealt with written skills and the 1979 report seems to address oral skills. Further, the 1979 report is subject of the same general problems already discussed for the 1976 report: no control for other factors where selection was not random and lack of information on the test used and how the analysis was done.

Finally, as noted earlier, there is a serious problem in the study with the number of students. Skutnabb-Kangas (1979, table 2) reports a sample of 687 immigrant students but the analysis shown in table 3-4 is based on only 177 students. The absence of data on more than three-quarters of the sample raises serious questions about the validity and generalizability of the results.

Webb County, Texas, Math Program

Description

Trevino (1968, 1970) looked at a bilingual program in a district surrounding Laredo, Texas, which is located on the Mexican border. The project was in one elementary school; the object of the program was to have each child learn a second language. English-speaking children consented to participate in the project. In the first and second grades a bilingual teacher taught in both languages. In the third grade one of the teachers was not bilingual, but because there were two third-grade classes the bilingual teacher taught in Spanish half the day and the monolingual English teacher taught in English half the day.

Trevino (1968) examined the effect of teaching math in the students' home language. A cohort of 183 bilingually taught students was compared with an earlier cohort of students taught in English (L2) in the first and third grades. Analysis was performed for all students and for a subset which had 3 years of school without retention. The pattern of results was the same for the total sample as for the subset. Of eight analyses of variance comparing the Spanish-speaking students in the bilingual program with the Spanish-speaking students who were taught only in English, four differences were significant in favor of the bilingual group. Basically, the bilingual group was superior on the arithmetic reasoning subtest, while no differences were demonstrated on the arithmetic fundamentals subtest.

Reasons for Rejection

There are three major problems with Trevino's study. First, Trevino made no attempt to match her treatment groups and earlier cohorts. To assume equivalence of cohorts within the same school is questionable. Trevino should have included statistical controls for cohort differences. Since she did not, it is possible the differences observed were nothing more than preexisting differences between the two cohorts and that the program is ineffective.

The second problem involves internal inconsistencies in Trevino's data. If the program worked, then the difference between the English-speaking and Spanish-speaking students should be less in the bilingual cohort than in the monolingual cohort taught entirely in English. This was not the case. Mexican-American children in the bilingual math class did as well as Mexican-American children in the English math curriculum when both groups were compared with English-speaking students in their respective cohorts.

A third problem with Trevino's study is the finding that the English-speaking students in the bilingual program did better in math than did English-speaking children in a regular English curriculum in previous years. There is little reason why a monolingual English-speaking child should do better in a Spanish curriculum than in an English-instructed math class.

The program was designed so that math was taught twice each day, once in Spanish and once in English. If this double teaching of math resulted in the bilingual cohort's having more total math instruction than the monolingual cohort did, then scores of both English- and Spanish-speaking students in the bilingual cohort would be higher. Interestingly, Trevino (1968) invokes this explanation to account for the superior performance of the bilingually taught Anglo children but fails to realize it can account for the performance of the Spanish speakers as well. We think these explanations can account for Trevino's results without invoking program success.

The Colorado Statewide Evaluation Study

Description

Egan and Goldsmith (1980) and Goldsmith (1980) report a statewide assessment of bilingual programs in Colorado for the 1979-80 school year. The authors used data from all available school districts in the State where gains in normal curve equivalents (NCES) could be determined for grades K through 4. (NCES are a type of standardized percentile score; the study is fundamentally a norm-referenced study.) The authors argue that since language-minority children would be expected to show a loss against the norm in the absence of treatment, program success is evidenced by classes showing either no change or an increase. They proceed to count such classes. In addition, they establish a second, stricter success criterion of a gain of at least seven NCES (one-third standard deviation).

Apparently, "no change" was defined as a posttest score within one-third standard deviation of the pretest score, so that some of the classes counted as evidence of success actually may have experienced declining performance. (Since these students were initially low-scoring, a decline of one-third standard deviation toward the tail of the distribution would cover a considerable range of scores.)

Goldsmith (1980) concludes that "overall, 87 percent of the program reported gains or maintenance of academic achievement...."

Reasons for Rejection

We reject Egan and Goldsmith's Colorado State study because their logic does not overcome the problem found in a norm-referenced design (see appendix B). The norm-referenced model assumes that the rate of improvement of students in the program would have been the same as that of the norming group in the absence of the special program. This may not be true with language-minority children.

A second major problem is that the statistical analysis does not support the conclusion. Technical readers will appreciate that the procedure leading to an 87-percent success rate cannot be taken as evidence of program success, given the regression toward the mean artifact in low-achieving populations and the authors' inclusion of a loss of up to one-third standard deviation in very low scoring children as evidence of success.

For nontechnical readers, the problem can be illustrated using some test data from a national sample of non-Hispanic students from the Sustaining Effects Study (SES). The percentile score distribution was divided into 20 parts—categories of 5 percentile points each—and the posttest percentile category was broken out by pretest percentile category. Therefore, the movement of students from fall to spring percentile categories can be counted. Since Egan and Goldsmith are dealing with low-achieving students, we limit the example to the four lowest categories in the fall (1st through 20th percentiles). Since Egan and Goldsmith counted losses of up to 7 NCES as no change, we will count a drop of 1 percentile category from fall to spring as no change. When we apply Egan and Goldsmith's logic to our data for regular students not in any special program, we find basically the same result as Egan and Goldsmith interpreted as a program effect. This is shown in table 3-7.

TABLE 3-7. PERCENTAGE OF STUDENTS' PERCENTILES CHANGING FROM FALL TO SPRING (READING)

Fall	Spring Percentile			
Percentile	Loss	No Change	Gain	No Change + Gain
1-5	0	36	64	100
6-10	0	51	49	100
11-15	15	33	52	85
16-20	22	32	46	78

Applying Egan and Goldsmith's logic to students not in special programs would lead to the conclusion that regular schooling produces impressive gains in low-achieving students—100 percent of those below the 11th percentile showed no change or a gain. Since this table is based on regular students not in any special program, it illustrates how the measurement error component of test scores could have accounted for Egan and Goldsmith's results.

Finally, the reporting of the results, especially in the executive summary (Goldsmith, 1980) is highly selective. For example, the author states, "In 1979, 75 percent of...kindergarten data...showed substantial gains in excess of seven NCES." This happens to be the most extreme positive cell out of their 12 (grade/time/year) table. The corresponding figure for second grade of the same year is only 35 percent.

The Santa Fe, New Mexico, Bilingual Program

Description

Leyba (1978) conducted a longitudinal study of a Spanish-English bilingual program in Santa Fe, New Mexico. Three elementary schools participated in a Title VII program for grades 1 through 6. Participants were volunteered by parents. The Mexican-American students participating were bilingual, although they were stronger in English than in Spanish. Anglo students also participated in the program. Three groups were used: a program longitudinal (continuous participation program) group, a nonlongitudinal group, and a nonrandom comparison group.

The Santa Fe program reported that the Title VII students showed an increasing capability in English language skills and mathematics over time. Leyba also states that in the majority of cases Title VII students outperformed the non-Title VII students in reading and mathematics. He also reports that the Title VII students over time surpassed or matched national norms in reading and math.

Reasons for Rejection

Leyba did not control for possible preexisting differences due to nonrandom selection. There is evidence that such differences existed. The program students had higher pretest scores than the comparison group in 51 of 63 cases. This strongly implies the program participants were better students to begin with and therefore might have made greater gains whether there was a program or not.

Much of the analysis is based on comparing grade-equivalent scores with national norms. For the reasons discussed in chapter 1 we find this an unacceptable analysis. However, in addition, raw score gains were tested for significance by the t-test. Ninety nonindependent t-tests were carried out; 15 were significant, 2 of the 15 in the wrong direction. This statistical analysis is doubtful. The nonindependence of the t-tests renders the probability level unknown so there is no justification for considering 15 significant results at the nominal 0.05 level to be evidence of a program effect.

Finally, Leyba hypothesizes a rank order for the three groups with the longitudinal participants better than nonlongitudinal participants, who, in turn, are better than the comparison group. No significant difference between the longitudinal and nonlongitudinal groups existed.

St. John Valley School District, Maine Bilingual Program

One of the most puzzling projects to assess is that of St. John Valley, which has been widely cited as an example of a successful bilingual program (Troike, 1978; Title VII, 1980). Located near the Maine border with Canada, the St. John Valley project was a French-English bilingual program Troike references three studies of the project (Dube and Herbert, 1975, Lambert et al., 1975; Veilleux, 1977). In our search, which included an ERIC search, personal contacts with Troike, Veilleux, and the St. John Valley School District, we were unable to obtain copies of two of these studies. We did, however, uncover two additional reports that were not included in Troike's bibliography; St. John Valley (1980) and American Institutes for Research (AIR, 1975e). The latter covers the first 5 years (1970-75) of the project's operation, while the former covers the last 5 years (1975-80). Neither of these reports cites Dube and Herbert (1975) or Veilleux (1977).

1. AIR (1975e)

Description

AIR (1975e) presents the results of local evaluations of the bilingual education program and supplements these results with some additional data it collected from the school districts. Using the locally developed data, AIR (1975e) presents a table of grade-equivalents in a norm-referenced comparison, finding that program students were performing at around grade level. In addition, in a supplemental analysis, AIR found that bilingual program students outperformed students in regular classes.

Reason for Rejection

The locally developed analyses were all based on a norm-referenced comparison and are, therefore, unacceptable. AIR's own analysis does attempt to create a comparison group composed of students in regular classrooms. However, the AIR analysis presents no data that demonstrate comparability between the program and comparison groups, presents no statistical adjustments for nonrandom assignment, and uses grade-equivalent scores.

Furthermore, AIR seems to misinterpret the data. AIR states, "Following single groups (i.e., reading diagonally downward), the results for each group get better, suggesting that the program may be refining its techniques and improving its achievements over time." AIR looked at cohorts longitudinally across several years (by reading the diagonals of the table). The referenced table presents both grade-equivalent scores and stanines. Because we find grade-equivalents inappropriate for this types of analysis, we will limit ourselves to looking at stanines. We calculated whether the stanine went up, went down, or stayed the same for each cohort from each year to the next. Scores went up in four instances, went down in five cases, and remained the same in seven cases. AIR's conclusion that the results of each group got better every year is not supported by the data.

2. St. John Valley (1980)

Description

The St. John Valley (1980) study looks at the Title VII program for 1975-1980. The report first presents data on the percentage of bilingual program students achieving specified curricular criteria.

In general, the program students met the specified learning goals. In addition, posttest data are analyzed in terms of grade-equivalent scores and stanines. Scores were around grade level and these results are presented as evidence of program success.

Reasons for Rejection

We reject both the criteria-referenced and norm-referenced analyses. These data, while valuable to the local school, cannot be used for our purposes. We have no means of determining how a control group would score on these criteria, and, therefore, cannot judge program effects.

Second, posttest scores are invalid for many of the same reasons. Although posttest grade-equivalent scores and stanines are shown for the program group, there was no control group. In the absence of control groups, the fact that the students scored somewhat above the national average cannot be interpreted as proof of program effectiveness.

3. Veilleux (1977)

Description

The most widely cited of the several evaluation reports on the St. John Valley program is that of Veilleux (1977). This study purports to show that achievement increased by comparing the results of the first 7 years of the program with test scores of students for the 3 years before the program started operation. His findings were these:

- Data gathered before bilingual project implementation and during the first 5 project years show general student achievement has improved since the inception of bilingual education for those students in the program.

- Project student achievement during the 7 years of bilingual education indicates that Title VII students in general, are achieving at or above national norms in reading, math, and language arts.
- Samplings in each of the participating districts indicate that Title VII bilingual education students are achieving higher scores on nationally standardized achievement tests than students who are not participating in the bilingual project.

Reasons for Rejection

There are a number of problems with Veilleux's analysis. His first conclusion is based on an analysis that compares different statistics from different metrics from different tests. The preprogram students were tested with the SAT test, while the program participants were tested with the MAT and SRA tests. The preprogram data are presented as the percentage of students below grade level. For the program students, the metric is the percentage of students below the average stanine score, which was five. Since the majority of preprogram students scored below the mean cutoff for their grade level while the majority of program students fell at or above the mean stanine score, the program was considered a success.

This is not an acceptable analysis since it is incorrect to compare a single score cutoff point with a wide-band category. Since the stanine scale divides the total distribution of scores into only nine categories, large numbers of students who would fall below a single score—that is, the percentile or grade-equivalent grade-level score will be included in the at-grade-level stanine (i.e., five).

Other tables in Veilleux make possible a more valid comparison. Table 3-8 extracts grade-equivalent scores from several of Veilleux's tables presenting postprogram data. Since 17 of the 23 (almost 80 percent) scores presented are "below grade level," it is clear that the shift from the grade-level metric to stanines drastically altered the perception of program success.

TABLE 3-8. POSTPROGRAM TEST SCORES

Grade	Grade Level	Reported Mean Grade-Equivalent
4	4.8	4.3, 4.7, 5.0, 4.8, 4.6, 4.8, 5.4, 4.9
5	5.8	5.2, 5.4, 5.4, 5.8, 5.4, 5.1
6	6.8	5.4, 5.3, 5.8, 4.8, 5.1, 5.3
7	7.8	6.3, 6.2, 6.8
8	8.8	NA

Moreover, the preprogram data included three school districts while the postprogram data came from only one of the three. It is instructive to compare the test scores for that one district for grade 6 only. Grade scores were available for both periods and are shown in table 3-9. Using the author's logic of comparing such scores, it is clear the program had harmful effects: students performed somewhat worse after 6 years of bilingual instruction than did their peers 7 years earlier who had not participate in a bilingual program.* Finally, cohort matching fails to fully control for differences that could affect performance in school. The author failed to provide evidence that the two groups were indeed comparable.

TABLE 3-9. GRADE 6 TEST SCORES

Preprogram SAT	Median Grade-Equivalent	Postprogram SRA	Mean Grade-Equivalent
Math Concepts	6.1	Reading	5.4
Math Computation	6.2	Math	5.3
Math Application	5.6	Language	5.8
Social Studies	5.8		
Science	5.3		

Veilleux's second finding is based on norm-referenced comparisons. As has been indicated before, a norm-referenced design is not appropriate for language-minority children.

Veilleux then presents data taken from AIR (1975a) comparing scores of an unspecified number of students from three grades in one of the school districts. This analysis has the following problems:

- Posttest-only design with nonrandom assignment.
- Use of grade-equivalents.
- No statistical tests of significance.
- The mean scores were "derived from quartiles and medians, and are approximate." In other words, there is no way of knowing

---

* Note that we do not necessarily believe this conclusion. We are only illustrating how the author's faulty analysis can be used to show exactly the opposite of what he claims. We think that the problems found in the analysis are so severe that nothing, either pro or con, can be concluded about the effects of the program. However, if the reader disagrees with our rejection of Veilleux's method, then the reader must also accept this negative evidence.

if the reported scores are anywhere near the students' correct scores.

Two tables summarizing data from one of the other schools again present a posttest-only comparison.

In conclusion, the Veilleux study is beset by a variety of methodological problems, and it fails to make a case for program success in St. John Valley.

4. Troike (1978)

Troike (1978) conducted a review of the recent literature including studies available on St. John Valley. We would have anticipated Troike's analysis of the program to closely parallel ours. However, it did not. The major discrepancy involved the use of a control group.

The description of the program given by Troike (1978) differs from descriptions in the other evaluations we reviewed. For example, different statements are made about the use of random assignment in St. John Valley. Troike states that "randomly selected bilingual-medium schools were matched with all English control schools having students of comparable IQ and socioeconomic status, and after 5 years (1970-75) bilingually-trained students were found to outperform students in the control schools in English language skills and math, and have continued to remain ahead ever since." In summarizing the results of the project evaluations AIR reviewed for the JDRP, AIR states that "other than national norms, no other comparison or control group was identified." In this report covering 1970-73, AIR developed its own control group data for three grades in one of the school districts (1974). In the St. John Valley evaluation covering 1975-80, there were no control group data. Therefore, when looking across all the reports it is not clear that there was either random assignment or for that matter, any kind of a control group for most of the 10 years covered by the project.

We also question Troike's interpretation of the results. In presenting the results of the study, Troike reports, "In 1969, prior to the beginning of the bilingual program, as many as 80 percent of students scored below grade level in language and math. Figures for 1974-75 show students in the program achieved average or above average stanines in all subject areas."* It is not clear this can be taken as indicative of a program effect, as the following illustration from the AIR data demonstrates.

The performance of the 1974 control group in the AIR study is compared with that of the bilingual program group. On average, the control students from three grades were 0.2 grade-equivalents below grade level with an average standard deviation of 0.9 grade-equivalents. Therefore, the control group scored at the national norm or at about the same level as participants in the bilingual program. Apparently, something happened between 1969 and 1974 to raise the performance of students not in the program as well as those in the program.

---

* AIR states, "Prior to 1970, studies made by guidance staff and supervisors in the area had shown that students in the schools were performing rather poorly." St. John Valley (1980) makes no reference to preprogram performance levels.

We cannot account for the differences between the studies Troika reviewed and those we read. However, we have more confidence in the validity of the AIR study prepared for the JDRP. These data would have been carefully reviewed by AIR and the final report closely checked for accuracy by both the Title VII program and the local school.

In conclusion, we are not sure what to interpret about what happened in St. John Valley. The study briefly described by Troika would seem to be strong evidence in favor of program success. Unfortunately, we have not been able to find such a study. Based on the two 5-year summary evaluations we read, we can only conclude that for the past 10 years students in the St. John Valley project have been achieving at right around the national norm. No evidence is given to suggest this was an accomplishment of the bilingual project.

Houston, Texas, Bilingual Program

Description

AIR (1975c) reports on Houston's Title VII bilingual education program during its fifth year of operation and presents data for years 3 through 5.

Although federally funded in part, the program receives most of its funding from State and local sources. The program schools are located in the eastern and northeastern parts of the city and have a student enrollment which is 53 percent Mexican-American, 42 percent black, and 5 percent Anglo, Asian, and American Indian.

Each bilingual classroom has a teacher and a half-time aide who assists with instruction. Instruction is blocked so that some time is devoted to Spanish reading and language arts. During the remainder of the day, instruction is in English for English-dominant and bilingual students, while students who are monolingual Spanish speakers receive additional subject matter instruction in Spanish after the lessons are presented in English.

Matched students from eight elementary schools with bilingual programs were compared with students from three schools without a program in operation over 3 years. The difference in gains between the two groups was tested with no pretest control included. However, 11 of 14 pretests differences favored the comparison group. Twelve of 14 tests of the significance of the difference in school year gains significiantly favored the program group. Additionally, six of eight tests of math gains significantly favored the program students.

Reasons for Rejection

Although a matching procedure was attempted, we reject the study because matching was not accomplished, leaving the study without adequate controls. AIR (1975c) noted the following limitations in the study, which we quote:

- Students were eliminated from the comparison groups in order to match mean pretest scores. This matching process introduced unknown biases into the data.
- The program had high attrition rates.
- All Spanish-dominant students in the comparison group were eliminated during the second year of the study when it was discovered they had been receiving ESL.

These last two problems render the matching process invalid. Students were matched during the first year of the study, but of the 124 first-grade participants in the first year of the study, only 38 were left in the fourth year. By eliminating the Spanish-dominant students from the control group in the second year of the study when they discovered all the Spanish-dominant controls were in ESL classes, the program managers introduced an important bias into the comparison. In addition, the study loses validity because of the failure to include an adjustment for pretest scores in the statistical analysis.

Harlandale, Texas, Bilingual Program*

Description

Harlandale Independent School District is in the San Antonio Greater Metropolitan area. Olesini (1971) decided to study Harlandale because of its composite bilingual population and its developed Spanish-English bilingual program. Olesini describes his sample as "being sixty third grade Mexican-American children" who were "selected at random." The treatment group had been in bilingual education classes for at least 2 years; the comparison group had always been in a regular English curriculum.

The children were compared on average gains in grade-equivalent scores by the t-test after the two groups had been found not to differ on IQ (verbal measure) or age. Scores of program participants and spelling and arithmetic computation were no different. However, the study found that program participants did better in vocabulary, reading, language, and arithmetic concepts.

---

* Harlandale is included in our total count of acceptable outcomes, table 2-1 and table 2-2. However, because of its use of grade-equivalents we have included it here rather than in chapter 2.

Reasons for Rejection

Contrary to his description, Olesini did not randomly select his program and comparison groups, although he did match the groups on IQ and age. His study is rejected, however, because of his use of grade-equivalent scores.

It is difficult to determine what happened even if grade-equivalents are accepted. On the pretest the entering third graders scored very high—3.7 for the program group and 3.5 for the comparison students. Neither group can be considered disadvantaged (as confirmed by their IQ scores, which were over 100). Over the 6 months separating the tests, the control group gained 3 months and the treatment group gained 7 months. The historic performance of the treatment group was better than 1 month gain for each month of instruction (see their pretest score). The same was the case for the comparison group. The comparison group had a bad year, since they had gained more than a month for each month of schooling for the first two grades, but in the third grade they gained only one-half month per month of school. The treatment group continued their historic pattern. This raises strong doubts that a program effect was demonstrated.

A second problem is found with the IQ control. If the students in the bilingual program are limited in their English performance as a result of their non-English-speaking background, their scores on a verbal IQ test (Otis Quick Scoring) must greatly underestimate their true potential. Since their IQ score averaged 103, they must be students of very high true ability (as reflected in their pretest scores).

Further, since there was no control for relative language dominance, the groups could have differed on relative language proficiency, yet had equal IQ scores. Differential true abilities, when exposed to differing amounts of English in the home, could have led to equal verbal abilities. Their school performance would then be different to the extent that learning in school is affected by true ability levels.

Alice, Texas Independent School District Bilingual Education Program

Description

Children entering kindergarten in Alice, Texas, were assigned to the Spanish-English bilingual programs on the basis of (1) test scores, (2) parental approval, and (3) space available. A nonrandom comparison group was formed by matching kindergarten language-dominance tests for project and nonproject classes. Since the original documents reviewed by AIR did not present any tests of statistical significance, AIR estimated significance using posttest standard deviations.

Out of 32 tests, AIR (1975a) found 17 that significantly favored the project students on the Inter-America test series. Twelve of the significant differences were in L1 tests, so only 5 of 16 tests of English performance favored the project students. Clearly the project did much better in teaching Spanish (which apparently was not taught in the comparison classes) than it did in English. However, on the 32 pretest scores given, the program group scored higher on 23, and the control group scored higher on 7, with equivalent pretest scores in only two groups. All but one of the significant gains favoring the project students occurred in grades where project students had an initial advantage at the start of the school year. Project students had higher pretest scores on 15 of the 16 Spanish tests and on 8 of the 16 English tests (1 more than the number by which the comparison group exceeded the control).

Reasons for Rejection

The authors made an attempt to match the treatment and comparison groups on the basis of class averages of students' scores on the "Oral Language Eligibility Test" upon entering kindergarten. Shortcomings in the procedure, however, introduced biases. Consider the kindergarten pretest scores for the 2 years covered by the report:

TABLE 3-10. ALICE, TEXAS, BILINGUAL PROGRAM
SCORES ON LANGUAGE-DOMINANCE TEST

	Spanish		English	
Year	Program	Comparison	Program	Comparison
1	57	42	48	49
2	55	48	48	52

Although the two groups look reasonably the same in their initial English scores, with the comparison students being slightly ahead, the treatment students show considerably better Spanish ability. Consider the implications suggested by the data that the two groups are equivalent in English but one does better in Spanish. The implication is that the program group had considerably better language skills overall. Therefore, no program effect needs to be postulated to account for why program participants (1) consistently had higher pretest scores in grades 1 through 3, and why they (2) gained more over the school year. This is exactly the pattern that would be expected from students with greater language abilities.

The bias identified in the pattern of the kindergarten matching is also found in the other grades. On all 7 Spanish reading pretests, the treatment group scored higher than the control group. The treatment group scored higher than the controls on 5 of 7 English reading pretest scores. Although the authors never tested the significance of pretest differences, we can test the pattern by the Binomial Test. The conclusions are that the treatment groups scored significantly higher in Spanish pretest scores (P=.008) and there was no difference in English (P=.227) across all grades, confirming the pattern discussed above for kindergarten.

Classroom turnover also distorts the matching procedure over time. Not only are students being lost over time (one comparison class dropped from 32 to 19 in one year), but students were also added to the study. In some grades sample size increased over the 2 years. The two cohorts experienced a 1-year loss of data of 22 percent and 34 percent, respectively. If these rates are indicative of the year to year turnover, the total turnover in the two cohorts not considered in the report would have been 88 percent and 170 percent, respectively. The effects of 5 years of turnover on the initial matching is unknown.

For the one grade where matching was unaffected by turnover, that is, the first year of participation for each cohort, there were no significant differences in the gains in English between the program participants and the controls.

Finally, we note that English scores favoring the control group occurred in the last year of the study in both cohorts. Therefore, there is no evidence of a cumulative impact.

English As A Second Language Study

Fairfax County, Virginia, English as a Second Language Program

Description

The Fairfax County, Virginia, public school system operates an English as a second language program in grades 2 through 12; most participants stay in this program for 1 to 3 years before being mainstreamed. Program students' gain scores on the California Achievement Test (CAT) varied widely in reading, vocabulary, reading comprehension, language mechanics, language expression, and spelling when measured (December 1980), but of the 71 comparisons made, 64 were educationally significant. Children ranged from the 7th to the 39th percentiles on the tests.

Former program children in grades 4 and 6 who were mainstreamed into the regular curriculum in Fairfax County by June 1977 were measured on the SRA and STEA in October 1977. Students in both grades scored highest in math (above the 50th percentile) and lowest in reading and science (28th to 38th percentile). A posttest 1 year later, again showed scores highest in math (63rd percentile) and lowest in reading (38th to 43rd percentiles). The mean composite score for math, social studies, reading, and science was about the 50th percentile.

The second group of fourth- and sixth-grade former ESL students, who left the program in June 1978, were tested the following October. Three of six content areas showed scores above the 50th-percentile level for both groups. The fourth graders performed almost at the 50th-percentile level for science, reading, and social studies, but the sixth graders showed lower achievement (reading, 39th percentile, and science, 43rd percentile).

Students in grades 4, 6, 8, and 11 categorized as speaking "mostly a language other than English" and having a "fairly good level" of English proficiency were tested on the SRA and STEA. Pre- and posttests indicated that, at all four grade levels, children scored highest in math (45th to 60th percentiles) and lowest in reading (13th to 45th percentiles). Grades were uniformly highest at grade 4 and lowest at grade 11.

The Fairfax County study is one of the few studies in the literature that addresses the issue of the difference between statistically and educationally significant gains. Given large enough sample sizes, very small differences will be found statistically significant. In many cases, these differences will be too small to make any practical difference. Therefore, a good evaluation practice is also to consider the question of whether any statistically significant difference is also big enough to be educationally important. Following Horst et al. (1975) and Linn (1978), Fairfax defined a gain to be educationally significant if it was greater than one-third standard deviation of the norm group. Fairfax found that 64 of 71 comparisons exceeded the criterion of educational significance. The effect of the Fairfax program on reading scores can be seen in table 3-11.

Reasons for Rejection

Although the Fairfax study contains much information of use to the local school system, in the final analysis the study addresses a question that is different from the issue of concern here. The Fairfax study looks at how well the bilingual students are performing in comparison with the national norm.

TABLE 3-11. CAT TOTAL READING PERCENTILE SCORES, FAIRFAX COUNTY ESL PROGRAM

Grade	Pretest	Posttest	Gain*	N
2	22	41	19	95
3	19	31	12	114
4	14	28	14	87
5	8	18	10	71
6	6	12	6	73
7	4	18	14	55
8	3	11	8	57
9	1	8	7	55
10	1	8	7	67
11	1	7	6	40
12	1	7	6	20

*The gain for each grade exceeded the educationally significant criterion of being larger than one-third standard deviation of the norming group.

We have discussed earlier the subtle but important difference between this question and the question of evaluating how well one method of instruction does compared with another. In the absence of a contreatment language-minority group, the Fairfax study contains no data relevant to the principal concerns of the present study.

By the standards generally used to assess the effectiveness of educational programs, these gains are very impressive. However, the interpretation of the gains is not so clear-cut. Note that although very large gains in percentile standing were made during just 1 year of the program, only students in grade 2 had an average percentile score above that given in the Proposed Language-Minority Rules as the cutoff point for bilingual education eligibility. From grade 3 on, although large gains were made, the final level of performance ranged from only the 7th to 18th percentiles.

Therefore, while the Fairfax County program is very successful when looked at in terms of gain over the school year, there are grounds to question whether students reached a level of performance at which their lack of English skills was no longer holding them back in school.

Other Inapplicable Studies

The preceding sections of this chapter discussed in some detail the problems we found with a number of studies that have regularly been cited by the proponents of bilingual education as supporting the effectiveness of TBE. We find the methodological/logical problems encountered in these studies to be so severe that we cannot accept them as valid evidence of program effects, for the reasons stated. In addition, we reviewed many other studies that we judged to be inapplicable. However, since these other studies have not been generally cited as evidence for TBE, it is not necessary to discuss them at length.

Table 3-12 summarizes the reasons for our rejection of all the studies we did not accept. Our reasons are summarized into eight categories, any one of which was sufficient to reject a study. The eight categories are:

1. No Adjustment. When students are not randomly assigned to treatment and control conditions, something (either matching or statistical adjustment) must be done to adjust for possible preexisting differences affecting learning that could bias the results of the tests. We rejected studies using nonrandom assignment where such adjustments were not made.

2. Gains Only. All students learn something over time. If a study demonstrates only that students' scores went up over the school year, no evidence of a program effect has been demonstrated. Gains have to be compared with scores of a proper control group of students not in the program in order to demonstrate program effects.

3. Norms. Since language-minority students do not develop in English, their second language, the same way monolingual English speakers develop in English, comparing the progress of language-minority students to norms based on monolingual English-speaking students is not acceptable.

4. Criterion Tests. Criterion-referenced tests can only be used for our purposes if there is universal agreement about the criterion or if a control group was included to demonstrate a program effect.

5. Statistics. Conclusions must be based on an appropriate statistical analysis. If no statistical analysis was done or if the analysis done was clearly contrary to generally accepted practice, we rejected the study.

6. Local Criteria. Some studies are designed to address specific, locally relevant criteria that do not apply to our concerns. We could not use such results since they speak to different questions.

7. GE. As discussed in chapter 1, we do not consider grade-equivalent scores to be an acceptable metric for program evaluation.

8. No Detail. If a study failed to provide enough information to enable us to be reasonably sure none of the above problems exists, we did not accept the study.

TABLE 3-12. METHODOLOGICAL REASONS FOR REJECTING STUDIES*

Study	No Adjustment	Gains Only	Norms	Criterion Tests	Statistics	Local Criteria	GE	No Detail	Other
AIR, 1975e (St. John Valley)	X	X	X		X		X		
AIR, 1975a (Alice, Tex.)	X							X	
AIR, 1975d (Philadelphia)	X							X	
AIR, 1975c (Houston)	X								
Alejandro, 1979	X							X	
Alston, 1980		X	X				X		
Alston, 1977						X			
Alston, 1980						X			
Alston, 1980			X			X	X		
Arce, 1979	X	X	X			X	X		
Arce and Sosa, 1975		X				X			
Battiste, et al., 1975								X	
Berget, 1980		X				X			
Birmingham, 1981		X			X			X	
Boyce, 1980	X								
Boyce, 1979	X								
Cahill and Foley, 1973		X	X						
Charlotte-Mecklenburg, 1980		X							
Cohen, B., 1971	X								
Collison, 1974									Author's interpretation flawed
Corpus Christi, 1980a			X		X				
Corpus Christi, 1980b			X		X				
Del Buono, 1971					X				
Development Associates, 1977								X	
Edwards, 1976								X	Literature review
Edwards and Smyth, 1976								X	Literature review
Egan and Goldsmith, 1981		X			X				
Elligett, 1980	X						X		
Ewanyshyn, 1978									Not a comparative evaluation, a study of L2 learning

(Continued)

TABLE 3-12. METHODOLOGICAL REASONS FOR REJECTING STUDIES* (Continued)

Study	No Adjustment	Gains Only	Norms	Criterion Tests	Statistics	Local Criteria	GE	No Detail	Other
Fairfax County, 1980			X						
Fort Worth, n.d.								X	
Ferris, 1979	X							X	
Garcia, 1978	X								
Genesee, 1976								X	Literature review
Ghini, 1979		X	X						
Giles, 1971								X	
Goldsmith, 1981			X		X				
Golub, 1981								X	
Goodrick, 1977		X							
Gudschinsky, 1971								X	Literature review
Guerrero, 1980		X							
Hall, 1970								X	
Harrison, 1980		X			X		X	X	
Herbert, 1971					X			X	
Horst, 1980									Author rejects 19 Bilingual Evals (PIP) as unsound
Illinois, 1981	X	X	X		X	X	X		
JDRP, 1977	X		X				X		
Kalmar, 1975					X		X	X	
Leyba, 1978	X				X				
McCarthy, 1976		X							
Macnamara, 1966						X			
Macnamara, 1970								X	Literature review
Miranda, 1979						X			Teacher "reports" of progress
Modiano, 1966	X								Interpretation of results not clear
Monzon, 1981	X							X	
Muller and Leonetti, 1970	X					X			
McCarthy, n.d.	X				X				

(Continued)

TABLE 3-12. METHODOLOGICAL REASONS FOR REJECTING STUDIES* (Continued)

Study	No Adjustment	Gains Only	Norms	Criterion Tests	Statistics	Local Criteria	CE	No Detail	Other
Offenburg, 1971	X			X					
Ojerinde, 1978									Good design; no findings available
Olesini, 1971							X		
Payne, 1973								X	
Payne and Medina, 1972								X	
Price, 1978		X							
Pryor, 1970								X	
Pryor, 1972								X	
Pryor, 1971								X	
Richard, 1979	X							X	
Rimm, 1975					X	X		X	
Rimm, 1978	X		X				X		
Rimm, 1980		X	X						
Rimm, 1977	X		X				X		
Rimm, 1979	X		X				X		
Rosier, 1977	X						X		
Rosier and Holm, 1980	X						X		
Rosier and Farella, 1976	X						X		
SEDL, 1979	X		X		X		X		
Schmid and Schonbein, 1980		X		X					
Scott, 1979						X			
Scudder, 1979									Compared 2 ESL programs
Skutnabb and Kangas, 1979	X				X			X	
Skutnabb-Kangas, and Toukomaa, 1976	X				X			X	
Smith and Smith, 1978		X				X			
South San Francisco, 1979		X		X			X	X	
St. John Valley, 1980			X		X		X		
Swain, 1978a								X	
Swain, 1978b								X	
Trevino, 1970		X	X				X		
Trevino, 1968	X								Confounded treatments

(Continued)

TABLE 3-12. METHODOLOGICAL REASONS FOR REJECTING STUDIES* (Continued)

Study	No Adjustment	Gains Only	Norms	Criterion Tests	Statistics	Local Criteria	GE	No Detail	Other
Valencia, 1971		X							
Valencia, 1970	X								
Veilleux, 1977	X		X		X		X		
Vorih and Rosier, 1978	X				X		X		
Willink, 1968									Compared 2 types ESL
Winter, 1979								X	
Yoloye, 1977									Interim report; no relevant results yet
Young, 1980		X	X				X		
Zimmer, 1976		X	X				X		

* We have not listed studies that do not address our questions.

CHAPTER 4

CONCLUSIONS

Our review of the literature exploring the effectiveness of bilingual education has examined two basic questions pertinent to the intent of current Federal policy:

1. Does transitional bilingual education lead to better performance in English?

2. Does transitional bilingual education lead to better performance in nonlanguage subject areas?

We examined well over 300 documents concerned with bilingual education. Only 28 studies were found to apply to our concerns and to meet our methodological criteria. These 28 methodologically sound studies included evaluations of pedagogical methods other than transitional bilingual education, namely English as a second language (ESL) and structured immersion. We present the results of our review and implications for Federal policy below.

Results

Of the several hundred studies covered by the review, only 28* were found to apply to our concerns and to meet our methodological criteria. Before discussing the studies we found to be methodologically acceptable, we should note that we found several studies that have previously been widely cited as evidence for the effectiveness of TBE to be methodologically unacceptable (Skutnabb-Kangas and Toukomaa, 1976; Skutnabb-Kangas, 1979; St. John Valley, 1980; Veilleux, 1977; Leyba, 1978; Trevino, 1968; Modiano, 1968; Egan and Goldsmith, 1981; Rosier and Holm, 1980; and AIR, 1975a, 1975c, 1975e).

Table 1 summarizes the 28 studies we found to apply to our two questions and to meet minimal methodological criteria; by comparison, Zappert and Cruz (1977) found 18 methodologically acceptable studies. For each study, table 1 gives the author, the grades of school encompassed, the number of students in the treatment and control groups combined, the languages used by the program, and the results the author(s) reported for second-language and math skills. The most frequent home language was Spanish, but a number of other languages were represented as well. The most common second language was English. In three studies, French was the second language. Most of the studies were neither longitudinal nor true experiments. Several studies included very large numbers of students.

---

* Includes the study by Olesini which used grade-equivalents.

For each study we examined, table 1 indicates whether the study was better than or equivalent to another approach. These comparisons were based on findings which were statistically significant. Some studies had mixed results, based either on tests or grade levels. Where mixed results are found, we have indicated the nature of the different results.

Structured immersion programs seem to have done particularly well.[10] Lambert and Tucker (1972) and Barik and Swain (1975) found second-language learning through structured immersion superior to ESL, and Pena-Hughes and Solis (1980) showed structured immersion superior to transitional bilingual education. As for nonlanguage subjects, Lambert and Tucker (1972), Barik et al. (1977), and Ramos et al. (1967) all showed that it is possible to teach math successfully in the second language. This finding suggests that if the curriculum is properly structured so that the means of communication is at a level the child can understand, there will be no negative consequences from teaching math in the second language. We found no data in these studies pertinent to other subject areas, which are often more dependent on verbal skills than math is. Ramos et al. (1967) reported the least favorable results for immersion in the literature. They found that immersion from grade 1 was as effective after 5 or 6 years as a TBE program in which all instruction was in L1 for grades 2 through 4, and in L2 thereafter.

The data on ESL instruction are not very informative. As just noted, two studies found structured immersion superior to ESL. Ames and Bicks (1978) and Balasubramonian et al. (1973) found that TBE programs which included an ESL component were no more effective than ESL alone. Lum (1971) had mixed results finding both that TBE programs which included an ESL component were no more effective than ESL alone and that ESL alone was superior to TBE. Legarreta (1979) found that a TBE program with ESL worked better than a TBE program without an ESL component.

Mixed findings were found for several of the studies. As a result, the reader will notice that there are more findings than there are studies. Mixed findings can be attributed to different achievement results either from grade to grade or between tests. Therefore, some studies may be counted more than once as showing a positive, no different, or negative finding.

With respect to TBE, positive outcomes pertaining to language performance were reported by Covey (1973), Carsrud and Curtis (1980), McConnell (1980), Olesini (1971), Plante (1976), Legarreta (1979), AIR (1975b), Cohen (1975), Kaufman (1968), and Zirkel (1972). However, the case for the effectiveness of TBE is called into question by studies that found no difference in second-language performance between treatment and comparison groups (Ramos et al. 1967; Ames and Bicks, 1978; Plante, 1976; Kaufman, 1968; Huzar, 1973; Legarreta, 1979; A. Cohen, 1975; SEDL, 1977; Carsrud and Curtis, 1980; Matthews, 1979; Skoczylas, 1972; McSpadden, 1979, 1980; Balasubramonian et al., 1973; Cottrell, 1971; Olesini, 1971; AIR, 1975b; Zirkel, 1972; Lum, 1971). Moreover, some studies found TBE to be less effective than either immersion or ESL (Lum, 1971; Pena-Hughes and Solis, 1980) and some found TBE to have negative effects by comparison with submersion (Danoff et al., 1977, 1978; Stern, 1975; Moore and Parr, 1978; A. Cohen, 1975; McSpadden, 1980).

Olesini (1971), A. Cohen (1975), and Ames and Bicks (1978) found that TBE improved acquisition of math skills. However, no effect was found by Danoff et al. (1978), Carsrud and Curtis (1980), Moore and Parr (1978), McSpadden (1979, 1980), A. Cohen (1975), Covey (1973), Olesini (1971), SEDL (1977), and Ramos (1967). Skoczylas (1972), McSpadden (1980), and Stern (1975) reported a negative effect.

Caution must be exercised in generalizing from table 1 because some issues of methodological adequacy remain. For example, Covey (1973) and McConnell (1980a, 1980b) report success for programs including TBE. However, the programs also included very low staff-student ratios--1 to 8 in the program studied by Covey (1980). Therefore, strong doubts exist as to whether the reported program effect was due to the use of bilingual instruction or to the small classes.

We also examined our findings to determine which studies would have been included if we loosened our criteria and accepted grade-equivalents. Only Olesini would then be included in our results. His results were generally favorable to TBE and have been included in table 1 and table 2.

It is instructive to look for patterns in the findings of all these studies. Table 2 summarizes our findings with respect to comparing alternative instructional approaches. We have grouped the 28 studies according to the comparisons they examine. Then, we have aggregated their findings according to whether the study had positive, no difference, or negative results in comparison to the other approach.* For example, the first comparison in table 2 looks at the effectiveness of TBE versus submersion. For second-language acquisition, 10 findings favored TBE, 15 findings found no differences between TBE and submersion, and 5 findings were actually negative for TBE.

The results in table 2 must be qualified. Rather than simply counting the number of studies with various outcomes, we must go beyond these tabulations and give more or less weight to different findings. For example, the study by Ames and Bicks (1978) (which found that TBE produced better math results than submersion did) took place in only one school district, while the Danoff et al. (1978) study (which found that TBE had no effect on math) was designed to be nationally representative. Therefore, Danoff's findings must be given considerably more weight. Nevertheless, a clear understanding of our finding can only be obtained by looking at the studies in the aggregate rather than looking at the studies in isolation. Our policy implications are presented below.

---

* Because, as already noted, some studies had mixed results, the reader will notice that there are more findings than there are studies. However, if a study administered five tests of which three had positive results and two negative ones, we would record only one positive and one negative result in our comparison tables.

Implications

We believe the literature makes a compelling case that special programs in schools can improve the achievement of language-minority children. There is no evidence, however, that a specific program should be either legislated or preferred by the Federal Government. Indeed, more research and demonstration projects with sound evaluation models are needed to determine which programs are effective with which types of children in which locations. The rest of this summary will present our findings.

Special Programs Can Improve Achievement in Language-Minority Students

The literature we reviewed indicates that special programs designed to overcome language difficulties in school can improve the achievement of language-minority children. The studies by Pena-Hughes and Solis (1980, 1981), Plante (1976), Huzar (1973), Covey (1973), Kaufman (1968), and Lum (1971) were true experiments, and all showed special programs to have positive or neutral effects. The ingenious nonexperimental design used by McConnell (1980a, 1980b) also seems to have firmly established the presence of a positive program effect. Positive effects also were reported in the nonexperimental studies of Zirkel (1972), Ames and Bicks (1968), AIR (1975b), Barik and Swain (1975), Olesini (1971), Barik et al. (1979), Lambert and Tucker (1972), Legarreta (1979), Carsrud and Curtis (1980), Cohen (1975), and Malherbe (1946). Note, though, that while special programs have been shown to be effective, this conclusion says nothing about the effects of any particular instructional approach.

The Federal Government Should Not Place Exclusive Reliance on Transitional Bilingual Education

For more than a decade, the Federal Government has worked toward institutionalizing transitional bilingual education as virtually the only approved method of instruction for language-minority children. TBE has been emphasized in Title VII funding decisions. TBE has been implemented nationwide by the Office for Civil Rights' interpretation of the Lau decision. And in 1980, the Department of Education proposed, with few exceptions, the legal mandate of transitional bilingual education through Federal regulations (a proposal that has been withdrawn by the current Administration).

When we reviewed the literature on the effectiveness of transitional bilingual education we did not find justification for such heavy reliance on this method of instruction. In order for the Federal Government to rely exclusively on one instructional method for meeting the needs of language-minority children, the following two conditions must hold:

1. There must be a strong case that the instructional method is uniformly effective.

2. Effective instructional alternatives should not exist. If the desired outcomes can be reached through more than one approach, the Federal Government should not constrain the options of local schools.

Only 28 studies that passed our methodological test addressed the effectiveness of TBE, and only 11 of the 25 studies looking at TBE reported a positive effect. Further, additional methodological problems in these studies impose strong limits on generalizing their results. Three studies suggest that the reported positive outcome could well have been due to other aspects of the program rather than to TBE itself (Covey, 1973; McConnell, 1980a, 1980b; Plante, 1976). In addition, a number of studies that used multiple-outcome measures found mixed results. Several other studies found a negative effect for TBE when compared with submersion, ESL, or immersion (Danoff et al., 1977; Moore and Parr, 1978; McSpadden, 1980; Skoczylas, 1972; Cohen, 1975; Lum, 1971; Stern, 1975; Pena-Hughes and Solis, 1980). Although we reviewed a limited number of immersion studies, each analysis of structured immersion generally found positive findings for that approach. Achievement in both language skill and subject matter knowledge was better through structured immersion than through ESL or TBE (Barik and Swain, 1975; Barik et al., 1977; Lambert and Tucker, 1972; Pena-Hughes and Solis, 1980).

These findings do not add up to a very impressive case for the effectiveness of transitional bilingual education. We conclude that TBE fails both tests for justifying reliance on it as the exclusive method for instructing language-minority children. There is no firm empirical evidence that TBE is uniquely effective in raising language-minority students' performance in English or in nonlanguage subject areas.

Since several States have followed the Federal lead in developing programs for language-minority children--in some cases, even legislating TBE--our analysis has implications beyond the Federal level.

Federal Policy Should Be Flexible

For more than a decade, Federal policy (as expressed through Title VII legislation, Title VII funding decisions, OCR implementation of the "*Lau* Remedies," and the August 5 Notice of Proposed Rulemaking) has emphasized transitional bilingual education to the virtual exclusion of alternative methods of instruction. We found through our analysis that this policy is not justified on the basis of educational effectiveness. While transitional bilingual education has been found to work in some settings, it has also been found ineffective and even harmful in other places. Furthermore, both of the major alternatives to TBE—structured immersion and ESL—have been found to work in some settings.

The commonsense observation that children should be taught in a language they understand does not necessarily lead to the conclusion they should be taught in their home language. They can be successfully taught in a second language if it is done right. The key to successful teaching in the second language seems to be to insure that the second language and subject matter are taught simultaneously so that subject content never gets ahead of language. Given the American setting, where the language-minority child must ultimately function in an English-speaking society, carefully conducted second-language instruction in all subjects may well be preferable to bilingual methods.

We conclude that it is very hard to say what kind of program will succeed in a particular school. Hence it seems that the only appropriate Federal policy is to allow schools to develop instructional programs that suit the unique needs and circumstances of their students.

There is no reason to assume a priori that the same approach that is applied to a rural Southwest Texas district with a large proportion of second-generation Hispanic children should also be applied to a district with a small group of Lao refugees in a Northern city. But Federal policy has been based on such an assumption over the years. Our review indicates that a fundamental change in Federal policy is needed.

We believe this change will require recognition by the Department of Education that other pedagogical methods for language-minority children can be effective and can meet civil rights criteria. Federal funding practices must encompass each of the special programs designed to meet the needs of language-minority children so that a more realistic balance among various program types is achieved.

A widespread structured immersion demonstration program is especially needed. Until now, the immersion method has been rejected on the basis of weak theoretical arguments.[11] Immersion may not transfer successfully from Canada to the United States, but this is an empirical question that should be answered by direct test. As a first step, the Department should immediately fund an extensive evaluation of the McAllen, Texas, program, which has a true experimental design for comparing the effectiveness of structured immersion and TBE for Mexican-American students of low socioeconomic status.

Given the complexity of the problem, it also seems that the Federal Government should provide the most current information on pedagogical methods for language-minority children so that school districts can make informed choices, adapting methods to their local needs.

Improved Bilingual Research and Program Evaluations Are Needed

More and better research and improved program evaluations in bilingual education are necessary if the needs of language-minority children are to be adequately met. The low quality of the methodology found throughout the literature is a serious problem. The major methodological problems with the literature include the following:

- The absence of random assignment between treatment and control groups,

- The use of study designs that cannot show a treatment effect in the absence of random assignment, such as the norm-referenced model or failure to use analysis of covariance, and
- The failure to apply appropriate statistical tests to demonstrate program effects.

These problems have particularly characterized Title VII evaluations. The Title VII bilingual program has begun to take steps to improve the quality of local results. However, our review has indicated that program evaluations are still of very poor quality; much improvement is still needed in this area.

Bilingual education involves many complex, difficult issues that have been little (or insufficiently) studied. Federal funding for research in the area of bilingual education was allotted for the first time under Part C of Title VII in 1978, with the Elementary and Secondary Education Act amendments (ESEA). The need for additional research is great.

Unfortunately, however, when Congress established the legislation in 1978, it limited research to examining transitional bilingual education specifically, rather than all pedagogical methods for students with limited English proficiency. As a result, Federal research has been skewed to focus on one method. Ultimately, the development of effective instructional programs for language-minority children will come about only through a more broadly based research agenda.

Areas for redirected research should include the following:

- A study of the divergent educational needs of language-minority children in the United States to include the examination of how these children's language deficiencies differ in their home language and English,
- Examination of the effectiveness of alternative instructional approaches and how these approaches meet the needs of different types of language-minority children,
- A reexamination of the theory of TBE (designed for monolingual L1 speakers), which may not be relevant to many of the language-minority students in the United States,
- Formulation of appropriate structured immersion curriculums,
- Examination of the methods of English as a second language (vocabulary drills versus meaningful English communication), and
- Examination of bilingual education teacher qualifications and the degree of fluency such teachers have in both languages.

TABLE 1

SUMMARY OF APPLICABLE STUDIES

					Languages[3]		Reported Results	
Author	Date	Grade	Design[1]	Number[2] of Students	L1	L2	L2	Math
AIR (Corpus Christi)	1975b	K-1	Longitudinal; analysis of covariance	393	Spanish	English	TBE no different from submersion in 1 grade; TBE better than submersion in 1 grade	
Ames and Bicks	1978	1-9	Analysis of covariance	669	Spanish and French	English	TBE no different from ESL alone	TBE better than ESL alone
Balasubramonian et al.	1973	K-3	Analysis of covariance and other adjustments	317	Spanish	English	TBE no different from ESL alone	
Barik and Swain	1975	K-2	Longitudinal; analysis of covariance	2,253	English	French	Immersion better than ESL	Math taught in L2 no different from math taught in L1
Barik et al.	1977	2-5	Longitudinal; analysis of covariance	*	English	French		Depending on year and grade, math taught in L2 was worse than, no different from, or even better than math taught in L1
Carsrud and Curtis	1980	4-5	Longitudinal; analysis of covariance	172	Spanish	English	TBE better than submersion in 1 grade; TBE no different from submersion in 1 grade	TBE no different from submersion

(Continued)

TABLE 1. SUMMARY OF APPLICABLE STUDIES (Continued)

Author	Date	Grade	Design[1]	Number[2] of Students	Languages[3]		Reported Results	
					L1	L2	L2	Math
A. Cohen	1975	K-3	Longitudinal analysis of covariance and other adjustments	90	Spanish	English	TBE no different from submersion on 86 of 100 language skills; submersion better than TBE on 11; TBE better than submersion on 3	TBE no different from submersion in 2 of 3 grades; TBE better than submersion in 1 grade.
Cottrell	1971	K-1	Analysis of covariance	470	Navajo	English	TBE no different from submersion	
Covey	1973	9	Random assignment	200	Spanish	English	TBE better than submersion	TBE no different from submersion
Danoff et al.	1977, 1978	2-6	Analysis of covariance and other adjustments; big study	8,900	Several	English	Submersion better than TBE	TBE no different from submersion
Huzar	1973	2-3	Random assignment; one-way analysis of covariance	160	Spanish	English	TBE no different from submersion	
Kaufman	1968	Junior High	Experiment; longitudinal	139	Spanish	English	TBE better than submersion on 2 component scores of a standardized achievement test and no different on 7 component scores in one school; TBE no different from submersion on 9 tests in another school.	
Lambert and Tucker	1972	1-4	Longitudinal; analysis of covariance	213	English	French		Math taught in L2 no different from math taught in L1

(Continued)

TABLE 1. SUMMARY OF APPLICABLE STUDIES (Continued)

					Languages[3]		Reported Results	
Author	Date	Grade	Design[1]	Number[2] of Students	L1	L2	L2	Math
Legarreta	1979	K	Analysis of covariance	80	Spanish	English	TBE better than submersion or TBE no different from submersion, depending on the test; TBE with ESL better than TBE without ESL component	
Lum	1971	1	Random assignment	55	Chinese	English	ESL alone better than TBE on 3 tests; ESL alone no different from TBE on 2 tests	
Matthews	1979	2,4, 6,8	Log-linear model	1,011	Many	English	TBE/ESL no different from submersion	
McConnell	1980	Pre-K -3	Longitudinal; subject as own control	1,020	Spanish	English	TBE better than submersion	
McSpadden	1979	K-1	Analysis of covariance	196	French	English	TBE no different from submersion	TBE no different from submersion
McSpadden	1980	K-2	Longitudinal; analysis of covariance	263	French	English	Submersion better than TBE in 1 of 3 grades; TBE no different from submersion in 2 grades	Submersion better than TBE in 1 of 2 graded; TBE no different from submersion in 2 grades
Moore and Parr	1978	K-2	Analysis of covariance	130	Spanish	English	Submersion better than TBE	TBE no different from submersion
Olesini**	1971	3	Matching	60	Spanish	English	TBE better than submersion in 1 of 3 components of a standardized test; TBE no different from submersion in one component	TBE better than submersion on 1 component of a standardized test; TBE no different from submersion on 1 component

(Continued)

TABLE 1. SUMMARY OF APPLICABLE STUDIES (Continued)

					Languages[3]		Reported Results	
Author	Date	Grade	Design[1]	Number[2] of Students	L1	L2	L2	Math
Pena-Hughes, and Solis	1980	K	Random assignment	156	Spanish	English	Immersion better than TBE	
Plante	1976	1-2	Longitudinal; experiment	72	Spanish	English	TBE better than submersion in 1 grade. TBE no different from submersion in 1 grade and for both grades combined	
Ramos et al.	1967	1-6	Longitudinal; matching	***	Hiligaynon Pilipino	English	TBE no different from immersion[4],****	TBE no different from immersion[4], ****
SEDL (Stebbins)	1977	K-3	Longitudinal; analysis of covariance and other adjustments; 5 sites	1,060	Spanish	English	TBE no different from submersion[4]	TBE no different from submersion
Skoczylas	1972	1	Analysis of covariance	47	Spanish	English	TBE no different from submersion	Submersion better than TBE
Stern	1975	4-6	Analysis of covariance	213	Spanish	English	Submersion better than TBE[4]	Submersion better than TBE[4]
Zirkel	1972	1-3	Matching; analysis of covariance	278	Spanish	English	TBE better than submersion on 1 test; TBE no different from submersion on 4 tests	

(Continued)

* Treatment = 73, control not given.

** Rejected for use of grade-equivalents only.

*** Unable to obtain information at present; however, the sample size was large.

****The classification of the instructional method used in this study cannot be determined, but our best guess is immersion.

[1] In the case of multiyear studies, the number of tested students was counted. Rather than counting the number of unique students, the study counted each year a student was tested as a separate instance.

[2] For studies not using random assignment, we note the method used to adjust for possible preexisting differences between the treatment and control groups. Analysis of covariance is a statistical method used to adjust for preexisting differences.

[3] L1 is the language-minority child's home language; L2 is the child's second language.

[4] This result represents our conclusion from the author's very complex analysis; see chapter 2 of the full report.

TABLE 2

SUMMARY OF FINDINGS FROM APPLICABLE STUDIES*

Transitional Bilingual Education Versus Submersion

TBE:	Second Language	Math
Positive	10	2
No Difference	15	9
Negative	5	3

Transitional Bilingual Education Versus English as a Second Language

TBE:	Second Language	Math
Positive	1	1
No Difference	3	NA
Negative	1	NA

Transitional Bilingual Education Versus Immersion*

TBE:	Second Language	Math
Positive	0	0
No Difference	1	1
Negative	1	0

Immersion Versus English as a Second Language*

IMMERSION:	Second Language	Math
Positive	1	NA

---

* Math scores found in immersion projects in Canada are difficult to compare with scores in regular English curriculums. What can be concluded, however, is that students can achieve equally well (or better) in math classes taught in L2 as in math classes taught in L1.

APPENDIX A

Appendix A

METHODOLOGY GLOSSARY

Age. The ease and rapidity with which children acquire a second language has been noted by parents, teachers, and linguists alike. The literature (Krashen, 1979) shows that, holding time and exposure constant--

- Learners who begin exposure to the second language (L2) during childhood generally achieve higher L2 proficiency than those beginning as adults.
- Older children acquire the early stages of language skills faster than younger children.
- Adults proceed through the early stages of language and sentence development faster than children.

Thus, adults and older children generally acquire second-language skills faster than younger children, but younger children will be language superior in the long run (Snow and Hoefnagel-Hoble, 1978).

Some recent research relates L2 development to Piaget's theory of cognitive phases (Kessler, 1980). Children learning a second language before age 6 have not moved from the stage of preoperational thought to concrete operational thought (Piaget and Inhelder, 1969), indicating a different means of acquiring L2 is found in older children. Asher and Garcia (1969) found that immigrant children who had arrived in the United States between the ages of 1 and 6 became almost native-like English speakers. Asher and Garcia also presented evidence that first starting to learn L2 between ages 7 and 10 has detrimental effects on both languages because sound discrimination recognition as well as pronunciation begin to deteriorate at the onset of puberty. Giles (1971) hypothesizes that L2 instruction during this age period causes interference between the two languages because L1 has not been fully developed (also see Anderson, 1978; Hughes, 1969).

Cognitive Ability. There is an extensive literature debating whether bilingualism has positive or negative cognitive effects (see Darcy, 1953; Peal and Lambert, 1962; Landry, 1974; Segalowitz, 1975; Humphrey, 1977; Coronado, 1979; Malherbe, 1945; Fishman, 1965; Jensen, 1963). This debate does not concern us except to note there are implications for the type of teaching methods used for bilingual children depending on the nature of the cognitive effects of bilingualism. However, these effects have yet to be well documented, so it would be premature to comment on them except to note that the verbal skills of bilingual children are generally poorer than their nonverbal skills.

More important for our purposes is the proposition that children differ in the speed with which they learn languages, whether due to differences in flair for languages or differences in general ability. If this is so, then more able children will acquire L2 more rapidly (see Johnson, 1953, cited in Albert and Obler, 1980). Humphrey (1977) compared nonverbal IQ scores for Hispanic children and found that the better the English language (L2) ability, the higher the nonverbal IQ. This finding is best interpreted as showing that mastery of a second language is easier for more intelligent children. (See also Ardal, cited in Cummins, 1980.) A related point is the learner's cognitive style (see Coronado, 1979, for a review).

Communities. Different communities have different characteristics that bear on the teaching of language minorities. Lambert and Sidoti (1980) point out that in every locale languages have a status dimension. In most U.S. communities languages other than English have low prestige. However, in those areas where English is a minority language or where minority community members are unafraid of losing their linguistic identity (Lambert and Tucker, 1972), bilingual education may well yield different results. Finally, communities differ in the number of minority groups present. Chicago, with over 100 languages, faces a problem different from that of a Southwest Texas community where only English and Spanish are typically found (for a review, see Skoczylas, 1973; Read, 1980).

Correlation. A correlation is a measure of the extent to which two things occur together. Daylight and the presence of the sun in the sky are perfectly correlated—you do not have one without the other. Most factors involved in education are not perfectly correlated. There is a high but imperfect correlation between achievement and IQ. Because the correlation is not perfect, if we want to understand why children differ in achievement, we must consider other relevant factors in addition to IQ.

All the relationships discussed in this report involve less than perfect correlations. Readers should keep in mind that everything said about bilingual education is prefaced with an unstated "there is a tendency for."

Ethnicity. The background nationality and sociocultural differences of the language minority child can affect learning English. Several authors have suggested that learning L2 is easier the more similar L2 is to L1 the home language, L1 (MacNamara, 1966). That is, it is easier for a Spanish-speaking child to learn English than it is for a Vietnamese-speaking child, because there are more linguistic similarities between English and Spanish than between English and Vietnamese. Cummins (1981) also postulates that the sociocultural factors of a language-minority child's background affect language learning. In a study of 1,200 immigrants in Toronto, Cummins postulates sociocultural differences as the reason that children of Chinese background learned English better and performed better academically than Franco-Ontarian children.

Gain Over the School Year. Several evaluations report only the difference between the program students' fall and spring scores, even testing the gain for significance. This procedure is unsound. Almost all students show some absolute gain over time, even if they are at the same time rapidly falling behind the norm.

Grade-Equivalent Scores. There are serious problems with grade-equivalent scores, as explained in "A Prototype Guide to Measuring Achievement Level and Program Impact on Achievement in Bilingual Projects" (Horst et al., 1980):

> They are based on the mistaken belief that a gain in test scores of one or more months for each month of instruction represents good progress. This is not true. Grade-equivalent scores provide an illusion of simplicity but, in fact, they are almost impossible to interpret, even for specialists in test construction. *Grade-equivalent scores should never be used by anyone for any purpose whatsoever*. (emphasis added)

The Hawthorne Effect. This is named after a famous study of industrial production which discovered that workers become more productive when they are the center of attention (i.e., being studied). The very fact, then, that the people in a study are studied can affect the study outcome.

Immigrant or Native-Born. Different problems face educators according to whether the student population with limited ability to speak English is foreign-born or native-born. Immigrants can be expected to be randomly distributed in age when they enter our schools. Therefore, the school is as likely to get a non-English-speaking immigrant student in the tenth grade as in the first grade. Native-born language-minority children are unlikely to make progress to the tenth grade without learning English. Therefore, the type of program operated by a school must change significantly with each increasing grade depending on whether the student population is immigrant or native-born.

Immigrant status has been linked to superior school performance by some researchers. Canadian research has shown that immigrants outperform the national norm in a number of areas once L2 is learned. Anecdotal observations by school personnel in the U.S. Southwest often conclude that recent Mexican immigrants outperform native-born Mexican-Americans (Carter, 1970; Troike, 1978; Cummins, 1981). The implication is that, although both native- and foreign-born children perform equally poorly in English, immigrants are more likely to have higher cognitive abilities or motivation. However, empirical evidence to support these observations is limited (Kimball, 1968; Anderson and Johnson, 1971).

Some studies have found that immigrants initially do less well than native-born language minorities in school. In Israel, a 15-year study by the Ministry of Education and Culture (1969) found that failure of immigrants from culturally different countries was caused by socioeconomic factors compounded by the character of Israeli schools, the structure of the educational system, and the curriculum, which was based on a different value system. Cardenas and Cardenas (1972), in their theory of incompatibilities, draw similar conclusions about the United States.

Baral (1979) found Mexican immigrants of lower socioeconomic status had lower academic achievement than Mexican-Americans of a higher status. However, Ferris (1979) found that junior-high children who immigrated from Mexico after grade 3 did as well in English written composition as native-born Mexican-Americans did. The foreign-born, moreover, were superior to Mexican-Americans despite a greater amount of travel time to Mexico, lower socioeconomic status, and more Spanish spoken in the home.

Lack of Standard Evaluation Format. Another problem that confounds effectiveness comparisons of programs stems from the lack of a standardized evaluation format. Alkin et al. (1974) examined this problem by looking at the Title VII experience in evaluation and the use of testing data. Swain (1978) has identified program testing as differing according to type, content, and utilization.

Local-Criterion-Referenced Designs. Communities may set up goals for their specific programs. such as parent involvement, curriculum development, and less absenteeism and then measure these goals through a criterion-referenced evaluation. This report is addresses a very specific Federal policy question dealing with student achievement aand English language acquisition. There may be studies that provide perfectly adequate answers to locally relevant evaluation questions that have no bearing on our concern here (for example, see Offenberg, 1970; Young, 1980; Goodrich, 1977). Therefore, we can make no use of these studies. Criterion-referenced tests are further discussed in the section on tests in this glossary.

Motivation/Self-Concept. It has been argued by some that many minority languages have low prestige in the United States. The majority, therefore, may view minority children as members of an inferior group, placing a negative value on the children, and depressing their self-concept, which in turn depresses school performance. De Avila and Ulibarri (1980) question this theory. Home language is identified with "significant others" who are very important in shaping the child's self-concept. The degree to which significant others are associated with reading and writing may be important to the development of self-concept and, consequently, literacy (Christian, 1976). Studies that show greater academic achievement and motivation by language minorities in classes with teachers of the same ethnicity illustrate this point (Modiano, 1973; Zirkel, 1972; von Maltitz, 1975).

Lum (1971) reviews several studies showing that students who accept the mainstream culture learn English more rapidly than do those who "cling to their own cultural group."

Closely related to self-concept is the role motivation plays in the child's school experience. Even though the literature is far from conclusive as to the role played by motivation, it is often put forth as a plausible explanation for discrepant findings. It is argued that the child's motivation to learn L2 and to otherwise perform in school is affected by (see Von Maltiz, 1975; Del Buono, 1971; Skoczylas, 1973; Rand, 1980) these factors:

- Whether L2 is the dominant language of the society or a highly valued second language,
- Whether the child learning L2 comes from the majority cultural group or from a minority group,
- The socioeconomic status of the family,

- The linguistic pattern of the community in which the child lives and the degree to which the community esteems L2 and L1,
- Parental language use,
- Parental attitudes toward L1 and L2,
- Subcultural differences, and
- Parental attitudes toward school in general.

The Novelty Effect. One interpretation of the research on educational improvement is that children respond to novelty. Following a change in the way a classroom is run, children will learn a little better. Once the new becomes commonplace, a change to something else again will succeed. Consequently, any innovative practice works when first tried, but loses its presumed effectiveness once it becomes the usual order of business.

Oral or Written Language Skills. Age seems to affect acquisition of oral and written skills differently. Cummins (1978, 1980) argues there is little transfer (facilitation) effect from learning oral L2 to learning written L2. Although there is no direct evidence supporting Cummins' theory, neither is there evidence refuting it. The implication is that no improvement in initially learning how to read (grades 1 and 2) either L1 or L2 will result from prior instruction (preschool or grade 1) in oral language skills in either language (see Fishman, 1965). Conversely, Kramer (1980), Venezky (1970), and Gudschinsky (1971) argue that oral skills are an important precursor to learning how to read. Goodman and Goodman (1978) report very different oral and written skill levels in language-minority children.

Parental Support. Lambert and Tucker (1972) emphasize that the St. Lambert program could not have been a success without the involvement of parents. Recognizing the importance of parental support, Title VII projects have emphasized the role of parent advisory groups (see also Del Buono, 1971).

Post hoc Explanations. The purpose of scientific study is to eliminate all possible explanations of an event except the one to be proved. If the study fails to support the chosen explanation, all the alternatives become viable. These after-the-fact alternative explanations are called post hoc explanations. They have very little value as scientific proof.

The bilingual literature has many examples of post hoc explanations for program results. Motivation, self-concept, and community differences are usually cited. By and large, the role these factors play in bilingual education has not been subjected to direct tests. That they have something to do with the outcome of a bilingual program is plausible, but not proven.

Posttest-Only Design. Some studies compare only posttest scores of students in the program and a nonrandomly selected comparison group. This approach is open to all the selection bias problems and the results of such studies cannot be relied upon.

Pretest Differences. An important characteristic of selection bias is that students in the comparison and program classes often differ on their pretest scores, which has serious consequences for the analysis of data. This problem is illustrated in the following data taken from Olesini's (1971) study of the Harlandale, Texas, bilingual program:

	Pretest Score	Posttest Score	Gain
Program.............	3.7	4.5	0.8
Comparison..........	3.4	3.6	0.2

Olesini found the difference between the gain of the program and comparison groups (0.6 grade-equivalents) was statistically significant and concluded the bilingual program was successful since students in the program learned more than students in the regular classes did. But the pretest scores show that students in the program were already 0.3 a grade-equivalent ahead before the program started. Their real gain over the comparison group during the program was 0.3, half of what Olesini reported.

As noted earlier, there are methods to compensate for selection bias problems introduced by nonrandom selection. This discussion illustrates how important it is that these controls be included in nonexperimental study designs, since without them it is difficult or impossible to know what would have been the performance of the program students in the absence of the special program. Depending on the exact combination of circumstances found at any particular school, the selection bias problem can make the program look either better or worse than it actually is.

The Pygmalion Effect. There is some evidence that student performance responds to the teacher's expectations of how well the student should perform. Therefore, if the teachers are convinced that a new program will improve the childrens' performance, or if the new program uses different teachers whose expectations are higher, the children will improve. This improvement has nothing to do with the new curriculum. It occurs because the teachers' attitudes have changed.

Relative Proficiency. This issue is rarely measured in the literature but is implied in the structure of bilingual programs. It suggests different treatments are appropriate depending on the degree of bilingualism of the child upon entering school. Battel et al. (1975) identify the degree of L1 dominance among students as one of the major factors affecting curriculum development in bilingual programs.

Relevant Factors. The most common reason for alternative explanations of results is that assignment to the program was not random. If the students selected to participate in a program differ from those selected to form the untreated comparison group on factors that affect learning, then the postprogram differences between the two groups could be due to the way they differed in their original disposition to learning rather than being due to an effect of the program.

One of the most important sources of alternative explanations for the results of a study is the complex nature of human behavior. Education is not a simple matter. Many factors affect a student's performance in school. If a study fails to consider the effects of these relevant factors on achievement, erroneous conclusions about the effectiveness of the program can result.

School performance due to a special program must be separated from other relevant factors. For example, it can be shown that test scores decline as the amount of Federal aid to a school district goes up. It is a mistake to conclude that Federal programs cause poor achievement because there is another relevant factor. Both low test scores and Federal aid are consequences of poverty. Poor schools have students who score low on tests and Federal aid is allocated to schools based on their poverty levels. A serious misinterpretation about the effect of Federal aid could be made if one ignored the relevant factor of poverty.

The first step in assessing the studies of effectiveness of programs for bilingual students is to determine what the relevant factors are so that the degree to which they were controlled for can be considered.

Interaction is a problem often found in association with relevant factors. Interaction means that the effect of one relevant factor changes when another relevant factor changes. There is, for example, an interaction between age and learning to write or speak a second language. It is well established that the younger the learners, the better they eventually come to speak L2. It is also well established that the older people are when they begin to learn how to read L2, the faster L2 is learned. Thus, there can be no single statement about the relationship between age and language learning.

There are numerous characteristics of language-minority children that affect their acquisition of English, including age, oral and written skills, parental support, cognitive ability, prior training in L2, ethnicity, self-concept, and motivation. Community and school attitudes also affect language learning and academic achievement. When measuring the effectiveness of bilingual education programs, one should somehow control for these relevant factors. If a study fails to consider these factors, its results may be open to serious question.

School Differences. While characteristics of language-minority children play a role in determining the outcome of an educational program, the educational program itself is obviously another source of factors affecting the course of learning. For example, McDonald and Elias (1976) of Educational Testing Service found that teacher performance makes a substantial contribution to what children learn. They found the second most important factor for predicting change, other than socioeconomic status, was what teachers did in the classroom. Research indicates that students with English proficiency have historically been provided fewer of the teacher practices related to student achievement (Dulay and Burt, 1979; Engle, 1975; Patniz et al., 1976). Teacher proficiency in L1 and teaching methods are of major importance to student success, as are teacher attitudes. Unfortunately, little is known about the art of attitude change.

However, de Kanter (1979) has employed a dissonance intervention strategy with teachers in Texas to address attitudinal and value inconsistencies which affect behavior, thus making teachers more open to students with English proficiency (also see Muller and Leonetti, 1970; Moore and Parr, 1978).

Basically, any dimension of the school setting related to the effectiveness of schooling can affect a bilingual program. For instance, a major problem in bilingual programs is the general lack of materials and qualified teachers (NTS, 1980). Separate program elements can be put together in different ways by different schools to form a bilingual program, and different combinations may be differentially effective.

Socioeconomic Status. Socioeconomic status (SES) has been shown to be at least two-thirds responsible for the relatively low achievement in school of Hispanics (Veltman, 1980). Moore (1978) has shown that students of higher socioeconomic status within bilingual programs do better than lower status groups. Rosenthal et al. (1981) found, in a nationally representative sample of elementary school students, that controlling for socioeconomic status accounted for most of the low achievement of Hispanic students relative to Anglo students. De Avila (1981) concluded that controlling for socioeconomic status reduced the total variance accounted for among Hispanic students, although it did not eliminate it. In inspecting National Assessment data we also found that cognitive performance of Hispanics closely resembled that of blacks and English-speaking low-SES groups, while neither blacks nor Hispanics resembled the higher SES Anglo/white students. Thus, home language may not be the primary cause of low achievement for many children from non-English-language backgrounds.

Tests. Evaluations use nationally normed referenced tests or criterion-referenced tests, or both. Standardized tests permit comparisons across programs, but they are not completely adequate measures because many of the groups tested differ considerably from the norming population. There is also the problem of cultural biases in standardized tests. Juarez (1974) pointed out that students do best on tests they take in their native tongue. However, Perez (1979) questioned court-ordered first-language testing in a study that found most of the language-minority children scored better in English. Lambert and Tucker (1972) found no difference in math score due to the language of the test.

Effectiveness studies measure different skills according to the program goals. Oral tests may measure factors completely different from reading, writing, and listening comprehension tests. Similarly, tests for cognitive skills, subject matter, and linguistic proficiency measure aspects of a program's effectiveness but cannot be compared to one another.

Although standardized achievement tests have problems, a better alternative does not usually exist, so we have given considerable weight to standardized achievement tests.

There are two problems with criterion-referenced tests from our perspective. First, if they are to be used in an evaluation of a program, an appropriate comparison group is absolutely essential. Second, the selection of the criteria presents a problem. They should be meaningful and appropriate to the question of interest; often they are not.

In order to use criterion-referenced tests for an evaluation of bilingual education across projects, we must specify the criteria to be assessed. This is impossible since we are reviewing existing studies which have already specified their locally relevant criteria. It should be noted that, although criterion-referenced test studies are of little use for our purposes, they may meet the evaluation needs of a local school. Alternatively, criterion-referenced test studies with control groups can be assessed across projects on the question of whether the program students outperformed the controls. In this case, all the usual cautions about selection bias apply. This is a moot point, however, as virtually no criterion-referenced test studies employing control groups exist.

Test scores have become the standard means of validating a project's success. Other data such as average daily attendance, parental involvement, and completion of secondary school may also be viewed as measures of project success, however. In fact, Paulston (1975) argues that the dropout rate is the best single measure of program success among language-minority secondary students.

APPENDIX B

BIBLIOGRAPHY

Abu-Sayf, F., and Herbolich, James B. "The Validation of Competence Areas of English as a Foreign Language Using Three Criteria." *Journal of Experimental Education*, 1979-80, 48(2), 170-73.

Adku, Sol. *Poverty Children and Their Language; Implications for Teaching*. New York: Grime and Stratton, 1979.

Afolayan, A. "An Adequate Theory of Bilingual Education for Africa." In Alatis, J.E., (ed.), *International Dimensions of Bilingual Education*. Washington, D.C.: Georgetown University Press, 1978.

Alatis, J.E. (ed.). *Current Issues in Bilingual Education*. Washington, D.C.: Georgetown University Press, 1980.

__________ . *International Dimensions of Bilingual Education*. Washington, D.C.: Georgetown University Press, 1978.

Alatis, J.E., and Tucker, G. Richard (eds.). *Language in Public Life*. Washington, D.C.: Georgetown University Press, 1979.

Albert, Martin L., and Obler, Loraine K. *The Bilingual Brain*. New York: Academic Press, 1978.

Alcala, Carlos M.; Rivera, Beatriz; and Thayer, Berta R. "Legal Significance of Lau." San Francisco: National Education Task Force De la Raza with Mexican American Legal Defense and Education Fund, Department of Justice, and Department of Health, Education, and Welfare, 1974.

Alejandro, Frank Z. "The Relationship of Bilingual Bicultural Education and Regular Education in the Verbal and Nonverbal Performance of Chicano Students." Unpublished doctoral dissertation, The Catholic University, Washington, D.C., 1979.

Alkin, Marvin et al. *Evaluation and Decision Making: The Title VII Experience*. CSE Monograph Series in Evaluation, No. 4. Los Angeles: University of California, 1974.

Alston, Herbert L., and Gavito, Alfredo A. *Houston Plan for Bilingual Education 1979-80 School Year Evaluation Report*. Houston: Houston Independent School District Research Evaluation and Accreditation, 1980.

______________________________ . "Houston Plan for Bilingual Education Program Review." December 1980. Mimeographed.

American Institutes for Research. "Alice Independent School District Bilingual Program, Alice, Texas." *Identification and Description of Exemplary Bilingual Education Programs*. Palo Alto, Calif.: Author, 1975a.

________________________. "Bilingual Education Program (Aprendemos En Dos Idiomas), Corpus Christi, Texas." Identification and Description of Exemplary Bilingual Education Programs. Palo Alto, Calif.: Author, 1975b.

________________________. "Bilingual Education Program (Title VII, P.L. 89-10), Houston, Texas." Identification and Description of Exemplary Bilingual Education Programs. Palo Alto, Calif.: Author, 1975c.

________________________. "Let's Be Amigos." Philadelphia, Pennsylvania." Identification and Description of Exemplary Bilingual Programs. Palo Alto, Calif.: Author, 1975d.

________________________. "St. John Valley Bilingual Education Project, St. John Valley, Maine." Identification and Description of Exemplary Bilingual Programs. Palo Alto, Calif.: Author, 1975c.

American Educational Research Journal. 1971, 8, 285-309.

Ames, J.S., and Bicks, Pat. "An Evaluation of Title VII Bilingual/ Bicultural Program, 1977-78 School Year, Final Report, Community School District 22." Brooklyn, New York: School District of New York, 1978. Mimeographed.

Anderson, James G., and Johnson, William H. "Stability and Change Among Three Generations of Mexican Americans: Factors Affecting Achievement."

Anderson, T. "Toward a Theory of Preschool Biliteracy." In Alatis, J.E. (ed.), International Dimensions of Bilingual Education. Washington, D.C.: Georgetown University Press, 1978.

Andersson, T., and Mackey, W. (eds.). Bilingualism in Early Childhood. Rowley, Mass.: Newbury House, 1972.

Andrews, Edward. "Montgomery County Public Schools' Program for Students with Limited English Proficiency." Montgomery County, Md., 1980. Mimeographed.

Arce, Aaron, and Sosa, Xavier. "ESEA Bilingual/Bicultural Project, 1974-1975 Final Evaluation Report." Austin, Tex.: Austin Independent School District, Office of Research and Evaluation, 1975.

Arce and Associates, Inc. "Project Rainbow. Pajaro Valley Unified School District. Watsonville, Calif.: 1978-79." Santa Ana, Calif.: Author, 1979. Mimeographed.

Arsenian, S. "Bilingualism in the Post-War World." Psychology Bulletin, 1975, 42(2), 65-86.

Asher, J., and Garcia, R. "The Optimal Age to Learn a Foreign Language." Modern Language Journal, 1969, 8, 334-41.

Baker, C. Personal Communication (via telephone), 1980.

Baker, K. "The Technological Fee." Paper presented at the Annual Meeting of the American Educational Research Association, Boston, April 1980.

Bamgbose, Ayo (ed.). Mother Tongue Education. London: Hodder and Stoughton, 1976.

Balasubramonian, K.; Seelye, H. Ned; and De Weffer, Rafaela Elizondo. "Do Bilingual Education Programs Inhibit English Language Achievement? A Report on an Illinois Experiment." Paper presented at the Seventh Annual Convention—Teachers of English to Speakers of Other Languages, San Juan, Puerto Rico, May 9-12, 1973.

Baral, David. P. "Academic Achievement of Recent Immigrants from Mexico." Nabe Journal, 1979, 111(3), 1-13.

Barik, Henri, and Swain, Merrill. "Three Year Evaluation of a Large Scale Early Grade French Immersion Program: The Ottawa Study." Language Learning, 1975, 25(1), 1-30.

Barik, Henri C.; Swain, Merrill; and Nwanunobi, Edna A. "English-French Bilingual Education: The Elgin Study Through Grade Five." The Canadian Modern Language Review, 1977, 33(4), 459-75.

Battiste, Marie et al. Study of Bilingual-Bicultural Projects Involving Native American, Indo-European, Asian and Pacific Language Groups. Palo Alto, Calif.: American Institutes for Research, 1975.

Bernal, Ernest M. "Longitudinal Evaluation of Bilingual Programs." FORUM, November 1980, III(9), 114.

Berke, I. Personal Communication, 1980.

Birman, Beatrice. "Problems of Overlap Between Title I and P.L. 94-142: Implications for the Federal Role in Education." Paper presented at the Annual Meeting of the American Educational Research Association, Boston, April 1980.

Birmingham Public Schools. "Bilingual Education Evaluation." Birmingham, Ala.: Author, 1980. Mimeographed.

Bloom, B.S. Human Characteristics and School Learning. New York: McGraw-Hill, 1976.

Boas, F. Handbook of American Indian Languages. Bureau of American Ethnology. Bulletin No. 40, pp. 1-83, 1911.

Bock, Geoffrey, and Stebbins, Linda B. Education as Experimentation: A Planned Variation Model Volume IV-B Effect of Follow Through Models. Cambridge, Mass.: ABT Associates, April 1977.

Borg, W. R. "Time and School Learning." In Denham, C., and Lieberman, A. (eds.), *Time To Learn*. National Institute of Education: Washington, D.C., 1980.

Bourdier, P. *Reproduction Culturelle et Reproduction Sociale*. Communication to the British Sociological Association, Durham, April 1970.

Boyce, Margaret. "Math Achievement of Participants and Non-Participants in the Bilingual Program." Paper presented at the Rocky Mountain Education Research Association Annual Meeting, Tucson, Ariz., September 26-28, 1979. Mimeographed.

______________. "Math Achievement of Participants and Non-Participants in the Bilingual Program." Paper presented at the Arizona Association for Bilingual Education, March 1980. Mimeographed.

Brembeck, Cole S., and Hill, Walker, H. *Cultural Challenges to Education*. Lexington, Mass.: D.C. Heath and Company, 1973.

Brisk, Maria Estela. "Language Policies in American Education." *Journal of Education*, Winter 1981, 163(1), 3-15.

Brown, Roger. *Social Psychology*. New York: The Free Press, 1965.

Buckingham, Thomas. "Appraising Bilingual Education: A Look Through Rose-Colored Glasses." *The Bilingual Review*, January-April 1980, VII(1), 72-74.

Burger, H.G. "Adapting Education Cross-Culturally." In Brembeck, C.S., and Hill, W.H. (eds.), *Cultural Challenges to Education*. Lexington, Mass.: Lexington Books, 1973.

Cahill, Robert J., and Foley, Joseph J. " Evaluation and Evaluative Research in an Urban Bilingual Program," 1973.

Cardenas, B., and Cardenas, J.A. "The Cardenas-Cardenas Theory of Incompatibilities." *Today's Education*, 1972.

Cardenas, Jose A. "The AIR Evaluation of the Impact of ESEA Title VII Spanish/English Bilingual Education Programs: An IDRA Response." San Antonio, Tex.: Intercultural Development Research Association, 1977. Mimeographed.

Carroll, J. "A Model for School Learning." *Teachers College Record*, 1963, 64, 723-33.

Carroll, J.B. *Language and Thought*. Englewood Cliffs, N.J.: Prentice-Hall, 1964.

Carsrud, Karen, and Curtis, John. "ESEA Title VII Bilingual Program: Final Report." Austin, Texas: Austin Independent School District, 1979, Mimeographed.

Carsrud, Karen Elizabeth. "Evaluation of Achievement Outcomes: Austin's Experience." Paper presented at the National Conference on Longitudinal Evaluation of Bilingual Programs, Austin, Tex., 1980.

Carter, L.F. The Sustaining Effects Study: An Interim Report. Santa Monica, Calif.: Systems Development Corporation, 1980.

Carter, Thomas P. "Mexican Americans in School: A History of Educational Neglect." New York: College Entrance Examination Board.

Center for Applied Linguistics. Bilingual Education: Current Prespectives, Social Science, Volume 1. Arlington, Va.: Author, 1977.

___________________________. Bilingual Education: Current Perspectives, Linguistics, Volume 2. Arlington, Va.: Author, 1977.

___________________________. Bilingual Education: Current Perspectives, Education, Volume 4. Arlington, Va.: Author, 1977.

Charlotte-Mecklenburg Schools. "Title VII Transitional Bilingual Education Summative Report 1979-80." Charlotte, N.C.: Author, 1980. Mimeographed.

Chicago Public School System. "An Analysis of the Proposed Rules for Non-Discrimination Under Programs Receiving Federal Financial Assistance Through the Education Department." Chicago: Author, Department of Multilingual Education, 1980. Mimeographed.

Christian, Chester, C. "Social and Psychological Implications of Bilingual Literacy." In Simoes, Antonio (ed.), The Bilingual Child. New York: Academic Press, 1976.

Chu-Chang, Mae. "The Dependency Relation Between Oral Language and Reading in Bilingual Education." Journal of Education. Winter 1981, 163(1), 30-55.

Cohen, A.A.; Boruck, M.; and Rodriquez-Brown, Flora. "Evaluating Evaluation." Bilingual Education Series 6. Arlington, Va.: Center for Applied Linguistics, 1979.

Cohen, Andrew D. A Sociolinguistic Approach to Bilingual Education. Rowley, Mass.: Newbury House Publishers, Inc., 1975.

Cohen, Bernard. "Final Evaluation Report of the 1970-71 New Haven Bilingual Education Program." Darien, Conn.: Dunlap and Associates, Inc., 1974.

Coleman, James; Hoffer, Thomas; and Kilgore, Sally. "Public and Private Schools." Paper presented at the Seminar on What Do We Know About Private Schools? Washington, D.C., April 7, 1981.

Collison, G.O. "Concept Formation in a Second Language: A Study of Ghanaian School Children." Harvard Education Review, 1974, 44(3), 441-57.

Contreras, Raul; Weibly, Gary; and Carsrud, Karen. Five Years of Bilingual Education: Austin Independent School District, Austin, Tex.: Office of Research and Evaluation, Austin Independent School District, 1980.

Cornejo, Ricardo J., and Cornejo, Luz O. Theories and Research on Second Language Acquisition. Las Cruzes, N.M.: Education Resources Information Center (ERIC), 1981.

Coronado, Leopoldo A. "The Effects of Differing Degrees of Bilingualism on the Cognitive Performance and Scholastic Achievement of Spanish/English Bilinguals." Unpublished doctoral dissertation, The University of Texas at Austin, 1979.

Corpus Christi Independent School District. "Evaluation Report for the Title VII Aprendemos en dos Idiomas Bilingual Program 1979-80." Corpus Christi, Tex.: Author, 1980.

______________________________ . "Follow Through Program Evaluation Report 1979-80." Corpus Christi, Tex.: Author, 1980.

Cottrell, Milford C. "Bilingual Education in San Juan County, Utah: A Cross-Cultural Emphasis." Paper presented at the American Educational Research Convention, New York City, February 1971.

Covey, Donald David. "An Analytical Study of Secondary Freshmen Bilingual Education and Its Effect on Academic Achievement and Attitude of Mexican American Students." Unpublished doctoral dissertation, Arizona State University, 1973.

____________________. Personal Communication (via telephone), 1981.

Culver City Unified School District. "El Marino Immersion Program." Culver City, Calif.: Author, 1977. Mimeographed.

Cummins, James. "The Construct of Language Proficiency in Bilingual Education." In Alatis, J.E. (ed.), Current Issues in Bilingual Education. Washington, D.C.: Georgetown University Press, 1980.

_____________ . "Educational Implications of Mother Tongue Maintenance in Minority Language Groups." Canadian Modern Language Review, 1978, 34(3), 395-416.

__________. "The Effectiveness of Bilingual Education." Seminar presented by the National Clearinghouse for Bilingual Education, Washington, D.C.: July 28, 1981.

__________. "The Entry and Exit Fallacy in Bilingual Education." NABE Journal, 1980.

__________. "The Influence of Bilingualism on Cognitive Growth: A Synthesis of Research Findings and Explanatory Hypotheses." Working Papers on Bilingualism, #9. Ontario, Canada: Ontario Institute for Studies in Education, 1976.

__________. "Linguistic Interdependence and the Educational Development of Bilingual Children." Review of Education Research, 1979, 2, 222-51.

__________. "Psycholinguistic Evidience." In Center for Applied Linguistics, Bilingual Education: Current Perspectives, Education, Volume 4. Arlington, Va.: Author, 1977.

Cummins, James, and Abdollell, Ali. "Adjustment of Immigrant Children in Bilingual Situations." Modern Journal of Applied Linguistics, 1980.

Cummins, Jim. "Empirical and Theoretical Underpinnings of Bilingual Education." Journal of Education, Winter 1981, 163(1), 16-29.

__________. "The Language and Culture Issue in the Education of Minority Language Children." Interchange, 1979-80, 10(1), 72-80.

__________. Personal Communication (via telephone), 1981.

__________. "Request to 'generally review both proposed bills currently being considered by the California Legislature for their consistency with current language development research'," California State Board of Education Memorandum, April 28, 1980. Mimeographed.

__________. "The Role of Primary Language Development in Promoting Educational Success for Language Minority Students." Paper prepared for the California State Department of Education Compendium on Bilingual-Bicultural Education.

Danoff, Malcolm N.; Coles, Gary J.; McLaughlin, Donald H.; and Reynolds, Dorothy J. Evaluation of the Impact of ESEA Title VII Spanish/English Bilingual Education Programs, Volume I: Study Design and Interim Findings. Palo Alto, Calif.: American Institutes for Research, 1977.

_______________ . Evaluation of the Impact of ESEA Title VII Spanish/ English Bilingual Education Programs, Volume II: Project Descriptions. Palo Alto, Calif.: American Institutes for Research, 1977.

_______________ . Evaluation of the Impact of ESEA Title VII Spanish/ English Bilingual Education Program, Volume III: Year Two Impact Data, Educational Process, and In-depth Analyses. Palo Alto, Calif.: American Institutes for Research, 1978.

_______________ . Evaluation of the Impact of ESEA Title VII Spanish/ English Bilingual Education Program, Volume IV: Overview for Study and Findings. Palo Alto, Calif.: American Institutes for Research, 1978.

Darcy, N.T. "A Review of the Literature on the Effects of Bilingualism Upon the Measurement of Intelligence." The Journal of Genetic Psychology, 1955, 82, 21-57.

Datta, Louis-ellin. "Review of Title VII Impact Study," HEW Memorandum, June 21, 1979. Mimeographed.

David, J.L., and Pelavin, S.H. "Evaluating Compensatory Education: Over What Period of Time Should Achievement be Measured?" Journal of Educational Measurement, 1978, 15, 91-99.

David, J.L., and Pelavin, S.H. Research on the Effectiveness of Compensatory Education Programs: A Reanalysis of Data. Menlo Park, Calif.: Stanford Research Institute, 1977.

Dean, Raymond S. "Comparison of PIAT Validity with Anglo and Mexican American Children." Paper presented at the Annual Convention of the Rocky Mountain Psychological Association, Phoenix, Arizona, May 1976.

De Avila, E.H. Relative Language Proficiency Types: A Comparison of Prevalence, Achievement Level, and Socio-Economic Status. Larkspur, Calif.: Delta Squared, 1981.

de Kanter, Ellen Ann Tharp. "Impact of a Dissonance-Based Intervention Strategy Upon Teachers' Attitudes Toward Equal Opportunity of Education for Mexican American Students." Unpublished doctoral dissertation, University of Houston, 1979.

_______________________ . Personal Communication (via telephone), 1980-81.

Del Buono, Xavier Antonio. "The Relationship of Bilingual/Bicultural Instruction to the Achievement and Self-Concept of Seventh Grade Mexican-American Students." Unpublished doctoral dissertation, Michigan State University, 1971.

Denton, Jon J., and McNamara, James F. "Conceptual Models for Determining the Influence of Classroom Teachers on Learner Cognitive Attainment." *Journal of Experimental Education*, 1979/80, 48(2), 165-69.

Development Associates, Inc. "Report on Extra-State Jurisdictions." Supporting Volume I of *A Study of State Programs in Bilingual Education* Washington, D.C.: Author, 1977.

________________________. "A Final Report on Their Status." *A Study of State Programs in Bilingual Education*, Washington, D.C.: Author, 1977.

________________________. "Inventory of Bilingual Education Provisions in State Legislation." Supporting Volume III of *A Study of State Programs in Bilingual Education*. Washington, D.C.: Author, 1977.

________________________. "Case Studies of Noteworthy Projects in Bilingual Education." Supporting Volume III of *A Study of State Programs in Bilingual Education*. Washington, D.C.: Author, 1977

Dubois, Betty Lou, and Valdes, Guadalupe. "Mexican-American Child Bilingualism: Double Deficit?" *The Bilingual Review*, January-April 1980, VII(1), 1-7.

Dulay, Heidi, and Burt, Marina. "Current State Level Trends in the Assessment of Language Minority Students: An Introduction." In Alatis, J.E. (ed.), *Current Issues in Bilingual Education*. Washington, D.C.: Georgetown University Press, 1980.

________________________. "The Efficacy of Bilingual Education." *Education Evaluation and Policy Analysis*, 1979, 1(5), 72-73.

________________________. "From Research to Method in Bilingual Education. In Alatis, J.E. (ed.), *International Dimensions of Bilingual Education*. Washington, D.C.: Georgetown University Press, 1978.

________________________. "The Relative Proficiency of Limited English Proficient Students." In Alatis, J.E. (ed.), *Current Issues in Bilingual Education*. Washington, D.C.: Georgetown University Press, 1980.

________________________. "Relative Proficiency Survey of Limited English Proficient Students." Unsolicited Proposal. San Francisco, Calif.: Bloomsbury West, 1979.

________________________. "Research Priorities in Bilingual Education." *Education Evaluation and Policy Analysis*, 1979, 1(3), 39-53.

Educational Management Services, Inc. An Evaluation of the 1975-76 ESEA Title VII Bilingual Bicultural Project, St. Paul Public Schools, Minneapolis, Minn.: Author, 1976.

Edwards, H.P., and Smythe, F. "Alternatives to Early Immersion Programs for the Acquisition of French as a Second Language." *The Canadian Modern Language Review*, 1976, 32(5), 524-33.

Edwards, Henry P. "Evaluation of the French Immersion Program Offered by the Ottawa Roman Catholic Separate School Board." *The Canadian Modern Language Review*, 1976, 33(2), 137-42.

Egan, Lawrence, and Goldsmith, Ross. "Bilingual Bicultural Education: The Colorado Success Story." 1980. Mimeographed.

Eisenhower, Esther. Personal Communication (via telephone), 1980.

Elligett, Jane K. "Bilingual Program Achievement Results at Sunset Hills and Tarpon Springs Elementary Schools, Pinellas County, Florida." Clearwater, Fla.: Pinellas County Schools, 1980.

Ember, Carol R. "Cross Cultural Cognitive Studies." *Annual Review of Anthropology*, 1977, 6, 33-56.

Endo, Todd. Personal Communication (via telephone), 1980.

Engle, Patricia Lee. "The Use of Vernacular Languages in Education: Language Medium in Early School Years for Minority Language Groups." *Bilingual Education Series #2*. Arlington, Va.: Center for Applied Linguistics, 1975.

Epstein, Noel. "Bilingual and Bicultural Education: The Role of the Scholar." In Alatis, J.E. (ed.), *International Dimensions of Bilingual Education*. Washington, D.C.: Georgetown University Press, 1978.

___________. *Language and Ethnicity: Policy Alternatives for Bilingual-Bicultural Education*. Washington, D.C.: Institute of Educational Leadership, George Washington University, 1978.

Ervin, S., and Osgood, C. "Second Language Learning and Bilingualism." In Osgood, C., and Sebeok, T. (eds.), *Psychologistics. Supplement to International Journal of American Linguistics*, 1954, 20(4), 139-46.

Ewanyshyn, Eugene. "Evaluation of a Ukranian-English Bilingual Program 1976-77." Alberta, Canada: Edmonton Catholic School, 1978.

Fairfax County Public Schools. "Evaluation Report: English as a Second Language Program." Fairfax, Va.: Author, Department of Instructional Services, 1980. Mimeographed.

Feldman, Carol, and Shen, Michael. *Some Language-Related Cognitive Advantages of Bilingual Five Year Olds*. Chicago, Ill.: Early Education Research Center, 1969.

Ferris, Malvyn Roger. "The Effects of the Early and Delayed Second Language Acquisition on the English Composition Skills of Spanish-Speaking Junior High School Students." Unpublished doctoral dissertation, Stanford University, 1979.

Fisher, C.W.; Berliner, D.C.; Filby, N.N.; Marilove, R.; Cohen, L.S.; and Dishow, M.M. "Teaching Behaviors, Academic Learning Time, and Student Achievement: An Overview." In Denham, C. and Lieberman, A. (eds.), Time To Learn. Washington, D.C.: National Institute of Education, 1980.

Fisher, Charles et al. Sample Description and Data Gathering Schedules: Part I of the Significant Bilingual Instructional Features Study. San Francisco: Far West Laboratory for Educational Research and Development, March 1981.

Fishman, J.A. "Bilingualism, Intelligence and Language Learning." Modern Language Journal, 1965, 49, 227-37.

Fishman, Joshua A. "The Social Science Perspective." In Center for Applied Linguistics, Bilingual Education: Current Perspectives, Social Science, Volume I. Arlington, Va.: Author, 1977.

Fort Worth Independent School District. "Effectiveness of the Bilingual Program in Fort Worth Schools." Fort Worth, Tex.: Author, 1975. Mimeographed.

Gaarder, A.B. "Bilingualism and Education." In Spolsky, B. (ed.), The Language Education of Minority Children. Rowley, Mass.: Newbury House, 1972.

Garcia, A.B. "Evaluation Report San Jose Area Bilingual Education Consortium, 1977-78." San Jose, Calif.: 1978.

Gearing, Frederick O, and Tindall, B. Allan. "Anthropological Studies of the Educational Process." Annual Review of Anthropology, 1973.

General Accounting Office. Comptroller General's Report to the Congress—Bilingual Education: An Unmet Need. Washington, D.C.: U.S. Department of Health, Education, and Welfare, 1976.

Genesee, Fred. "The Suitability of Immersion Programs for All Children." Canadian Modern Language Review, 1976, 494-515.

Ghini, C. "Terrebonne Parish French-English Bilingual Education Project Year End Interim Evaluation Report." New Orleans: G&A Ltd., 1979. Mimeographed.

Giles, W.H. "Cultural Contrasts in English/French Bilingual Instruction in the Early Grades." Paper presented at the Conference on Child Language, Chicago, Ill. November 1971.

Glass, Gene V. "Statement on Federal Regulation of Education for English Deficient Pupils." Boulder, Colo.: University of Colorado, Laboratory of Educational Research, 1980. Mimeographed.

Goldsmith, R.P. "Analysis of the Final Evaluation Reports Submitted to the Bilingual Unit of the Colorado Department of Education by Bilingual Education Programs Funded Under the Colorado Bilingual/Bicultural Education Act During the 1979-80 School Year." Boulder, Colo.: BUENO, 1981.

Golub, Lester S. "Literacy Development of Bilingual Children." *Bilingual Journal*, Spring 1981, v(3), 9-14.

Gonzalez, Carlos. "Effects of Ethnic Background, Socioeconomic Status, Education, General Mental Ability, English Proficiency, and Age on Aptitude Test Scores." Unpublished doctoral dissertation, New Mexico State University, 1977.

Goodman, K.S., and Goodman, Y.M. "Reading of American Children Whose Language Is a Stable Rural Dialect of English or a Language Other Than English: Final Report." Detroit: Wayne State University, 1978.

Goodrick, G. Ken. "Final Evaluation Report Title I ESEA Bilingual Project 1976-77." Houston, Tex.: Houston Independent School District, Program Planning, Research, and Evaluation, August 1977.

Goor, Jeanette; Baker, Keith; and Farris, Elizabeth. *ESEA Title I Evaluation: School District Needs for Technical Assistance 1979.* Washington, D.C.: National Center for Education Statistics, 1980.

Gray, Tracy. "Challenge to USOE Final Evaluation of the Impact of ESEA Title VII Spanish/English Bilingual Education Programs." Arlington, Va.: Center for Applied Linguistics, 1977. Mimeographed.

Gray, Tracy; Convery, H. Suzanne; and Fox, Katherine M. "The Current Status of Bilingual Education Legislation." *Bilingual Education Series: 9*. Washington, D.C.: Center for Applied Linguistics, 1981.

Gudschinsky, S.C. "Literacy in the Mother Tongue and Second Language Learning." Paper presented at the Conference on Child Language, Chicago, Ill., November 1971.

Guerrero, F.J. "Final Evaluation of the 1979-80 State Grant Bilingual Education Program." New Haven, Conn: New Haven Public Schools, 1980.

Hall, R. "Learning To Read in Two Languages: Statements from the Research Literature on Reading in Bilingual Programs." Philadelphia, Pa.: Philadelphia School District, 1970.

Hamilton, Don. "Evaluation Instruments for Bilingual Education. A Revision of Tests in Use in Title VII Bilingual Education Projects." Unpublished doctoral dissertation, 1972.

Hansen, Donald A., and Johnson, Vicky A. Language Attitudes of Parents, Students, and Teachers in Bilingual Classrooms. Berkeley, Calif.: University of California, 1981.

Harrison, Helene. "Final Evaluation Report for the State Bilingual Education Program of the Harlandale Independent School District." Harlandale, Tex., 1980. Mimeographed.

Hemenway, J.A.; Ming-Mei, Wang; Kenoyer, C.E.; Hoepfner, R.; Bear, M.B.; and Smith, G. The Measures and Variables in the Sustaining Effects Study, Report #9, Santa Monica, Calif.: System Development Corporation, 1979.

Herbert, Charles H., Jr. "Initial Reading in Spanish for Bilinguals." Paper presented at the Conference on Child Language, Chicago, Ill., 1971.

Hernandez-Chavez, E. "Language Maintenance, Bilingual Education and the Philosophies of Bilingualism in the U.S." In Alatis, J.E. (ed.), International Dimensions of Bilingual Education. Washington, D.C.: Georgetown University Press, 1978.

Heyneman, Stephen P. "Why Impoverished Children Do Well in Uganda Schools." Comparative Education, 1979, 15(2), 175-85.

Horn, Thomas D. "Three Methods in Developing Reading Readiness in Spanish Speaking Children in First Grade." The Reading Teacher, 1966, 20, 38-42.

Horst, D.P.; Johnson, D.M.; Nava, H.G.; Douglas, D.E.; Friendly, L.D.; and Roberts, A.O.H. "Prototype Evaluation Manual." An Evaluation of Project Information Packages (PIPs) as Used for the Diffusion of Bilingual Projects, Volume III. Mountain View, Calif.: RMC Corporation, 1980.

______________. "Summary Report." An Evaluation of Project Information Packages (PIPs) As Used for the Diffusion of Bilingual Projects, Volume I. Mountain View, Calif.: RMC Corporation, 1980.

______________. "Technical Discussion and Appendices." An Evaluation of Project Information Packages (PIPs) As Used for the Diffusion of Bilingual Projects, Volume II. Mountain View, Calif.: RMC Corporation, 1980.

Humphrey, Ronald J. "The Effects of Language Dominance as Determined by the Shutt Primary Language Indicator Test on the Measurement of the Intellectual Abilities and Achievement Ratings of Mexican-American Children in K, 1, 2 Grades." Unpublished doctoral dissertation, Northern Arizona University, 1977.

Huzar, Helen. "The Effects of an English-Spanish Primary Grade Reading Program on Second and Third Grade Students." Unpublished M. Ed. thesis, Rutgers University, 1973.

Illinois State Board of Education. The First Annual Program Summary and Evaluation Report of Transitional Bilingual Education Programs in Illinois. 1979-1980. May 1981.

"Immersion Schools C'est Bon For Students." In Education USA, December 1, 1980, p. 10.

Inter-America Research Associates. The Third Annual Report of the National Advisory Council on Bilingual Education. Washington, D.C.: Author, 1977.

Israel Ministry of Education and Culture. "Education and Culture." Paper presented at the Symposium "Integration of Immigrants from Different Countries of Origin in Israel," Jerusalem, October 1966.

Jensen, J. Vernon. "Effects of Childhood Bilingualism, I." Elementary English, 1962a, 39, 132-143.

Jensen, J. Vernon. "Effects of Childhood Bilingualism, II." Elementary English, 1962b, 39, 358-366.

Joint Dissemination Review Panel. "Follow Through: The Corpus Christi Independent School District 1976-1977." Washington, D.C.: U.S. Department of Health, Education, and Welfare, 1977. Mimeographed.

______________________________. Tulare Follow Through Project 1976-1977, Austin, Tex.: Southwestern Education Development Laboratory, 1977. Mimeographed.

Joint Legislative Audit Committee, Office of the Auditor General, State of California. Implementation of Bilingual Education in California. Sacramento, Calif.: Author, 1980.

Juarez, Manuela, and Raimondis, Nina Mendoza. Effects of the Language of the Instructions on Test Performance: California Achievement Test. San Antonio, Tex.: Edgewood Independent School District, 1974.

Kalmar, R.R. "The Culver City Schools' Spanish Immersion Program: Low Cost/High Yield." CATESOL Occasional Papers #2. 1975.

Kaufman, Maurice. "Will Instruction in Reading Spanish Affect Ability in Reading English?" Journal of Reading, 1968, 17, 521-27.

Kessler, C., and Quinn, M.E. "Positive Effects of Bilingualism on Science Problem-Solving Abilities." In Alatis, J.E. (ed.), Current Issues in Bilingual Education. Washington, D.C.: Georgetown University Press, 1980.

Kessler, Carolyn. Personal Communication (via telephone), 1981.

Kessler, Carolyn, and Quinn, Mary Ellen. "Emergence of Semantic Notions in Child Bilingualism." Paper presented at the Los Angeles Second Language Research Forum, Los Angeles, October 1978.

Kimball, William L. "Parental and Family Influences on Academic Acievement Among Mexican-American Students." Unpublished doctoral dissertation, University of California at Los Angeles, 1968.

Klibanoff, L., and Haggart, S. Summer Growth and the Effectiveness of Summer School. Santa Monica, Calif.: RMC Research Corporation, 1981.

Kramer, V.R. "The Effect of Auditory Discrimination Training of Spanish Speaking Children on Auditory Discrimination and Sound Letter Associations." In Outstanding Dissertations in Bilingual Education. Rosslyn, Va.: National Clearinghouse for Bilingual Education, 1980.

Krashen, Stephen. "The Input Hypothesis." In Alatis, J.E. (ed.), Current Issues in Bilingual Education. Washington, D.C.: Georgetown University Press, 1980.

Krashen, Stephen; Long, Michael A.; and Scarcella, Robin C. "Age, Rate, and Eventual Attainment in Second Language Acquisition." TESOL Quarterly, 1979, 13(4), 573-82.

Kuhn, Sara, and Long, Huey B. "Relationship of Formal Operations and Syntactic Complexity in Oral Language of Adult Women." Journal of Research and Development in Education, 1980, 13(3), 52-65.

Kirp, David. "Law, Politics, and Equal Education Opportunity. The Limits of Judicial Involvement."

Lambert, W.E. "Cognitive and Socio-Cultural Consequences of Bilingualism." Paper presented at a Multidisciplinary Conference on Bilingualism held on the Plattsborgh Campus of the State University of New York, March 12-13, 1976.

Lambert, W.E., and Sidoti, N. "The Selection of Appropriate Languages of Instruction and the Use of Radio for Education in Less Developed Countries." Washington, D.C.: World Bank, 1980. Mimeographed.

Lambert, W.E., and Tucker, G.R. "Bilingual Education of Children: The St. Lambert Experience." Rowley, Mass.: Newbury House, 1972.

Landry, Richard G. "A Comparison of Second Language Learners and Monolinguals on Divergent Thinking Tasks at the Elementary Level." Modern Language Journal, 1974, 58, 10-15.

Lange, Dale L., and Clifford, Ray T. "Testing in Foreign Languages, ESL, and Bilingual Education 1966-1979: A Select Annotated ERIC Bibliography." Language in Education: Theory and Practice. Arlington, Va.: Center for Applied Linguistics, 1980.

Larsos, Mildred L., and Davis, Patrica M. (eds.). *Bilingual Education: An Experience in Peruvian Amazonia*. Washington, D.C.: Center for Applied Linguistics, 1981.

Laumann, L.F. *Effect of Project Head Start, Summer 1965: A Second Look at the Equality of Economic Opportunity Study*. Institute for Research on Poverty, University of Wisconsin, Madison, 1969.

Legarreta, Dorothy. "The Effects of Program Models on Language Acquisition by Spanish Speaking Children." *TESOL Quarterly*, 1979, 13(4), 521-34.

________________. "Language Choice in Bilingual Classrooms." *TESOL Quarterly*, 1977, 11(1), 9-16.

Lenneberg, E.H. "Cognition and Ethnolinguistics." *Language*, 29, 463-71, 1953.

Leyba, Charles F. "Longitudinal Study Title VII Bilingual Program Santa Fe Public Schools." Los Angeles, Calif.: California State University, 1978.

Liberty Union High School District. "Final Evaluation Report 1979-80 Title VII." Brentwood, Calif.: Author, 1980. Mimeographed.

Lum, John Bernard. "An Effectiveness Study of English as a Second Language (ESL) and Chinese Bilingual Methods." Unpublished doctoral dissertation, University of California at Berkeley, 1971.

Macauley, Juliet Iyabode. "Bilingual Education in Nigeria. The IFE YORUBA Project." Paper presented at the Conference on Foreign Language Education and Bilingual Education. Arlington, Va.: Center for Applied Linguistics, 1979.

MacCarthy, Karen. "English as a Second Language Program Evaluation." White Plains, N.Y.: White Plains Public Schools, Office of Research, Testing, and Evaluation, 1976. Mimeographed.

Mackey, W. "A Typology of Bilingual Education." In Anderson, T., and Boyer, Mildred (eds.), *Bilingual Schooling in the United States*. Washington, D.C.: U.S. Government Printing Office, 1970, Vol. 2, 63-82.

Macmillan, Deborah L. "Language Policies for African Primary Education: Summary of the Anglophone Research Literature." Washington, D.C.: World Bank, 1980. Mimeographed.

Macnamara, J. "Comparative Studies of Reading and Problem Solving in Two Languages." Paper presented at the Fourth Annual TESOL Convention, San Francisco, Calif., March 1970.

Macnamara, John. *Bilingualism and Primary Education*. Edinburgh: Edinburgh University Press, 1966.

Malherbe, E.C. The Bilingual School. London: Longmans Green, 1946.

Matthews, T. "An Investigation of the Effects of Background Characteristics and Special Language Services on the Reading Achievement and English Fluency of Bilingual Students." Seattle, Wash.: Seattle Public Schools, Department of Planning Research and Evaluation, 1979.

McConnell, Beverly. "Individualized Bilingual Instruction. Final Evaluation, 1978-79 Program." Pullman, Wash.: 1980.

________________ . "Individualized Bilingual Instruction for Migrants." Paper presented at the International Congress for Individual Instruction, Windsor, Ontario, October 1980.

McConnell, Beverly Brown. "Effectiveness of Individualized Bilingual Instruction for Migrant Students." Unpublished doctoral dissertation, Washington State University, 1980.

McDonald, Frederick J., and Elias, Patricia. "Beginning Teacher Evaluation Study, Phase II, 1973-74 Executive Summary Report." Princeton, N.J.: Educational Testing Service, 1976.

McGuire, Raymond. "Testimony Concerning the Proposed Lau Regulations." Denver, Colo.: Aurora Public Schools, September 1980. Mimeographed.

McLaughlin, Barry. "Language Learning in Bilingual Instruction Literature Review. University of California at Santa Cruz.

McSpadden, J.R. "Acadiana Bilingual Bicultural Education Program. Interim Evaluation Report 1978-79." Lafayette Parish, La.: 1979.

____________ . "Acadiana Bilingual Bicultural Education Program. Interim Evaluation Report 1979-80." Lafayette Parish, La.: 1980.

Mayeske, G.W.; Okada, T.; Cohen, W.; Beaton, A. Jr.; and Wisler, C. A Study of the Achievement of our Nation's Students. Washington, D.C.: U.S. Government Printing Office, 1973.

Miller, G.A., and McNeil, D. "Psycholinguistics," in Lindzey, G., and Aronson, E. (eds.), The Handbook of Social Psychology, 2nd ed. Reading, Mass.: Addison-Wesley, 1969.

L. Miranda and Associates, Inc. An Evaluation of the State Funded Illinois Transitional Bilingual Education Program Final Report. Washington, D.C.: Author, 1979.

Modiano, Nancy. "A Comparative Study of Two Approaches to the Teaching of Reading in the National Language." Unpublished doctoral dissertation, New York University, 1966.

____________ . Indian Education in the Chiapas Highlands. New York: Holt, Rinehart and Winston, 1973.

Moore, F.B., and Parr, G.D. "Models of Bilingual Education: Comparisons of Effectiveness." The Elementary School Journal, 1978, 79(2), 93-97.

Muller, Douglas G., and Leonetti, Robert. "A Cumulative Summary of the Three Years of the Sustained Primary Program for Bilingual Students, 1967-70." Las Cruces, N.Mex.: Division of Plans and Supplementary Centers, BESE, 1970.

National Advisory Council for Bilingual Education. "The Prospects for Bilingual Education in the Nation." Fifth Annual Report. 1980-81.

National Assessment of Educational Progress. Hispanic Student Achievement in Five Learning Areas: 1971-75. Washington, D.C.: U.S. Government Printing Office, 1977.

National Association of Bilingual Education. "A Compilation of Response to the Evaluation of the Impact of ESEA Title VII Bilingual Education Program Conducted by the American Institutes of Research (AIR)." Fourth Annual Report of the National Association of Bilingual Education.

National Center for Education Statistics. "Hispanic Student Achievement in Five Learning Areas 1971-75." National Assessment of Educational Progress. Washington, D.C.: Author, 1977.

National Clearinghouse for Bilingual Education. Outstanding Dissertations in Bilingual Education. Rosslyn, Va.: Author, 1980.

National Institute of Education. A Compendium of Bilingual Education and Related Projects. Washington, D.C.: U.S. Department of Education, 1980.

______________________________. "A Re-examination of the Evaluation of the Impact of ESEA Title VII Spanish/English Bilingual Education Programs." Washington, D.C: Author, 1979. Mimeographed.

Nelsen, Francois, and Fernandez, Roberto M. Achievement of Hispanic Students in U.S. Schools. Chicago, Ill.: The National Opinion Research Center, 1981.

Nickel, K.N. "Experimentation, Extrapolation, Exaggeration--Thy Name is Research." 1979.

NTS Research Corporation. "Technical Proposal: Assisting the Office of Bilingual Education in Meetings Its Capacity Building Mandate." Durham, N.C.: Author, 1980.

Offenberg, Robert M. "Let's Be Amigos: Title VII Bilingual Project Evaluation of the First Year, 1969-70." Philadelphia, Pa.: Philadelphia Public Schools, 1970.

________________. "Let's Be Amigos: Title VII Bilingual Project Evaluation 1970-71." Philadelphia, Pa.: Philadelphia Public Schools, 1972.

Ojerinde, Adedibu. "The Use of a Mother Tongue, Yoruba, as a Medium of Instruction in Nigerian Schools." Unpublished doctoral dissertation, Cornell University, 1978.

Olesini, J. "The Effect of Bilingual Instruction on the Achievement of Elementary Pupils." Unpublished doctoral dissertation, East Texas State University, 1971.

Oller, J.W. "The Language Factor in the Evaluation of Bilingual Education." In Alatis, J.E. (ed.), *International Dimensions of Bilingual Education*. Washington, D.C.: Georgetown University Press, 1978.

Oller, John W., Jr. "A Language Factor Deeper than Speech: More Data and Theory for Bilingual Assessment." In Alatis, J.E. (ed.), *Current Issues in Bilingual Education*. Washington, D.C.: Georgetown University Press, 1980.

O'Malley, J. Michael. "Preliminary Report from the Children's English and Services Study (CESS)." HEW Memorandum, December 29, 1978. Mimeographed.

O'Malley, M. "Review of the Evaluation of the Impact of ESEA Title VII Spanish/ English Bilingual Education Program." *Bilingual Resources*, 1978, 1(2), 6-10.

Oren, Ditza L. "Cognitive Advantages of Bilingual Children Related to Labeling Ability." *Journal of Educational Research*, January-February 1981, 74(3), 163-69.

Orland, Martin E., and Reisner, Elizabeth. *Building Local Capacity for Bilingual Education Through ESEA Title VII: Introducing the Issues*. Durham, N.C.: NTS Research Corporation, 1980.

Parker, W. *The National Interest and Foreign Language*. Washington, D.C.: U.S. Department of State, 1962.

Paulston, C.B. "Ethnic Relations and Bilingual Education: Accounting for Contradictory Data." In Troike, R.C., and Modiano, N. (eds.), *Proceedings of the First Inter-American Conference on Bilingual Education*. Arlington, Va.: Center for Applied Linguistics, 1975.

Paulston, Christina Bratt. "Rationales for Bilingual Education Reforms: A Comparative Assessment." *Comparative Education Review*, 1978, 22, 402-19.

__________________. "Research." In Center for Applied Linguistics, *Bilingual Education: Current Perspectives, Linguistics, Volume 2*. Arlington, Va.: Author, 1977.

Paviola, Ema Teresa Josefa. "The Development and Measurement of Syntactic and Morphological Variables in the Written Spanish of Native Spanish-Speaking Students in Fourth to Ninth Grades." Unpublished doctoral dissertation, Pennsylvania State University, 1979.

Ojerinde, Adedibu. "The Use of a Mother Tongue, Yoruba, as a Medium of Instruction in Nigerian Schools." Unpublished doctoral dissertation, Cornell University, 1978.

Olesini, J. "The Effect of Bilingual Instruction on the Achievement of Elementary Pupils." Unpublished doctoral dissertation, East Texas State University, 1971.

Oller, J.W. "The Language Factor in the Evaluation of Bilingual Education." In Alatis, J.E. (ed.), International Dimensions of Bilingual Education. Washington, D.C.: Georgetown University Press, 1978.

Oller, John W., Jr. "A Language Factor Deeper than Speech: More Data and Theory for Bilingual Assessment." In Alatis, J.E. (ed.), Current Issues in Bilingual Education. Washington, D.C.: Georgetown University Press, 1980.

O'Malley, J. Michael. "Preliminary Report from the Children's English and Services Study (CESS)." HEW Memorandum, December 29, 1978. Mimeographed.

O'Malley, M. "Review of the Evaluation of the Impact of ESEA Title VII Spanish/ English Bilingual Education Program." Bilingual Resources, 1978, 1(2), 6-10.

Oren, Ditza L. "Cognitive Advantages of Bilingual Children Related to Labeling Ability." Journal of Educational Research, January-February 1981, 74(3), 163-69.

Orland, Martin E., and Reisner, Elizabeth. Building Local Capacity for Bilingual Education Through ESEA Title VII: Introducing the Issues. Durham, N.C.: NTS Research Corporation, 1980.

Parker, W. The National Interest and Foreign Language. Washington, D.C.: U.S. Department of State, 1962.

Paulston, C.B. "Ethnic Relations and Bilingual Education: Accounting for Contradictory Data." In Troike, R.C., and Modiano, N. (eds.), Proceedings of the First Inter-American Conference on Bilingual Education. Arlington, Va.: Center for Applied Linguistics, 1975.

Paulston, Christina Bratt. "Rationales for Bilingual Education Reforms: A Comparative Assessment." Comparative Education Review, 1978, 22, 402-19.

____________________. "Research." In Center for Applied Linguistics, Bilingual Education: Current Perspectives, Linguistics, Volume 2. Arlington, Va.: Author, 1977.

Paviola, Ema Teresa Josefa. "The Development and Measurement of Syntactic and Morphological Variables in the Written Spanish of Native Spanish-Speaking Students in Fourth to Ninth Grades." Unpublished doctoral dissertation, Pennsylvania State University, 1979.

Payne, I.V. "Southeastern New Mexico Bilingual Program. Program Accomplishment Audit, 1972-73." Artesia, N. Mex., 1973.

Peal, E., and Lambert, W. "The Relation of Bilingualism to Intelligence." *Psychological Monographs*, 1976, 75(27), 1-23.

Pelavin, S.; Bryson, S.; Burnett; and Peterson, S. *A Study of Year-Round Schools: Final Report*. Menlo Park, Calif.: SRI International, 1979.

Pena-Hughes, Eva. Personal Communication, 1981.

Pena-Hughes, Eva, and Solis, Juan. "abcs." McAllen, Tex.: McAllen Independent School District, 1980. Mimeographed.

Penalosa, Fernando. *Chicano Sociologists: A Brief Introduction*. Rowley, Mass.: Newbury House, 1980.

Penalosa-Stromquist, Nelly. "Teaching Effectiveness and Student Achievement in Reading in Spanish." *The Bilingual Review*, May-August 1980, VII(1), 95 104.

Penfield, W., and Roberts, C. *Speech and Brain Mechanisms*. Princeton, N.J.: Princeton University Press, 1959.

Perez, F.M. "The Performance of Bilingual Children in the Spanish Standardized Illinois Test of Psychologistic Abilities." Unpublished doctoral dissertation, University of Illinois, 1979.

Pettigrew, T.J., and Pajonas, P.J. "The Social Psychology of Heterogenous Schools." In Brembeck, C.S., and Hill, W.H. (eds.), *Cultural Challenges to Education*. Lexington, Mass.: Lexington Books, 1973.

Piaget, Jean. "Piaget's Theory." In Musem (ed.), *Carmichael's Manual of Child Psychology, Third Edition*. Vol. 1. New York: Wiley, 1970.

Piaget, Jean, and Inhelder, Barbel. *The Psychology of the Child*. New York: Basic Books, Inc., 1969.

Pifer, Alan. "Bilingual Education and the Hispanic Challenge." New York: Carnegie Corporation, 1979.

Plante, Alexander J. *A Study of the Effectiveness of the Connecticut "Pairing" Model of Bilingual-Bicultural Education*. Hamden, Conn.: Connecticut Staff Development Cooperative, 1976.

Price, Aida. "Evaluation Report I.S. 184 Demonstration Project in Bilingual/Bicultural Education. Title VII ESEA District 7 Bronx 1977-78." New York: New York City Board of Education, 1978. Mimeographed.

Pryor, Guy C. "Final Educational Accomplishment Audit of the Bilingual Education Program. Harlandale Independent School District, 1969-70." San Marcos: Southwest Texas State University, 1970.

__________ . "Final Educational Accomplishment Audit of the Bilingual Education Program. Harlandale Independent School District, 1970-71." San Marcos: Southwest Texas State University, 1971.

__________ . "Final Educational Accomplishment Audit of the Bilingual Education Program. Harlandale Independent School District 1971-72." San Marcos: Southwest Texas State University, 1972.

Pulte, William. "Are Bilingual-Bicultural Programs Socially Divisive?" *Integrated Education*, 1978, 16, 31-33.

Ramirez, M.; Macaulay, R.K.S.; Gonzales, A.; Cox, B.; and Perez, M. "Spanish English Bilingual Education in the U.S.: Current Issues, Resources and Research Priorities." In *Bilingual Education Series: 6*. Arlington, Va.: Center for Applied Linguistics, 1977.

Ramos, M.; Aguilar, J.V., and Sibayan, B.F. *The Determination and tation of Language Policy*. Philippine Center for Language Study Monograph Series 2. Quezon City, The Philippines: Alemor/Phoenix, 1967.

Read, J.S. "A Sociolinguistic Study of Lau Language Maintenance." *Outstanding Dissertations in Bilingual Education*. Rosslyn, Va.: National Clearinghouse for Bilingual Education, 1980.

Regulatory Analysis Review Group. "Report of the Regulatory Analysis Review Group Submitted by the Council on Wage and Price Stability. Department of Education's Proposed Title VI Language Minority Rules for Bilingual Education." Washington, D.C.: Council on Wage and Price Stability, October 1980. Mimeographed.

Richard, L.L. "Oral Language Gains in a French-English Bilingual Education Program in Iberia Parish (La.) as Measured by a Repetition Instrument." Unpublished doctoral dissertation, University of Texas at Austin, 1979.

Riegehaupt, Barbara. "El Marino School: Aquise habla espanol." *Valley News*, June 3, 1980.

Rimm, S. "Evaluation of the 1976-77 ESEA Title VII Bilingual/Bicultural Project, St. Paul Public Schools." Watertown, Wis.: Education Assessment Services, Inc. 1977.

______ . "Evaluation of the 1977-78 ESEA Title VII Bilingual/Bicultural Project, St. Paul Public Schools." Watertown, Wis.: Education Assessment Service, Inc., 1978.

______ . "Evaluation of the 1978-79 ESEA Title VII Bilingual/Bicultural Project, St. Paul Public Schools." Watertown, Wis.: Education Assessment Service, Inc., 1979.

______ . "Outcome Evaluation of the Title VII Bilingual/Bicultural Program for Laotian/Hmong, St. Paul Public Schools. Final Report." Watertown, Wis.: Education Assessment Service, Inc., 1980.

Rimm, S., et al. "Process Evaluation Report of the Bilingual/Bicultural Program of Education ESEA Title VII St. Paul Public Schools." Watertown, Wis.: Education Assessment Service, Inc., 1975.

Rogers, Donald D.; Flynn, Donald L.; Errecart, Michael T.; and Hoff, Jeanette M. A Cost Analysis of Follow Through Projects. Bethesda, Md.: RMC Research Corporation, April 1977.

Rosenthal, A.S.; Milne, A.; Ginsburg, A.; and Baker, K. "A Comparison of the Effects of Language Background and Socioeconomic Status on Achievement Among Elementary School Students." Washington, D.C: Applied Urbanetics, Inc., 1981. Unpublished mimeograph.

Rosier, Paul Webb. "A Comparative Study of Two Approaches of Introducing Initial Reading to Navajo Children. The Direct Method and the Native Language." Unpublished doctoral dissertation. Northern Arizona University, 1977.

Rosier, Paul, and Farella, M. "Bilingual Education At Rock Point--Some Early Results." TESOL Quarterly, 1976, 10(4), 379-88.

Rosier, Paul, and Holm, Wayne. "The Rock Point Experience: A Longitudinal Study of a Navajo School Program (Saad Naaki Bee Na'nitin)." Bilingual Education Series #8. Arlington, Va.: Center for Applied Linguistics, 1980.

Rossi, Peter H. "Comments on Title VII Evaluation." Amherst, Mass.: University of Massachusetts, 1979. Mimeographed.

Royster, Eugene C. Language Immersion Programs: Implications for Bilingualism. Durham, N.C.: NTS Research Corporation, 1979.

Salviani, Candida. "A Comparison of the Acquisition of Piagetian Conservation Concepts and the Attainment of Mathematical Skills by Bilingually Educated Bilingual Second Grade Children and Monolingually Educated Bilingual Second Grade Children." Unpublished doctoral dissertation, New York University, 1977.

Sanchez, Juanita. "A Comparison of Achievement of Mexican and Mexican-American Children in the Areas of Reading and Mathematics When Taught Within a Cooperative and Competitive Goal Structure." Unpublished doctoral dissertation, University of California at Santa Barbara, 1979.

Schmid-Schonbein, Gisela. "Evaluation of the Teaching of English to German Children of Preschool Age." English Language Teaching Journal, 1980, 34(3), 173-79.

Scott, Eloise. "Buffalo Public Schools Spanish-English Program Evaluation, 1975-77." Buffalo, N.Y.: Buffalo Public Schools, 1979.

Scudder, B.E.T. "A Comparative Study of the Effects of the Use of a Diagnositc/Prescriptive Approach Versus a Tutorial Approach to the Teaching of English to Non-English Speaking Elementary School Children in a Large Urban School District in Colorado." Unpublished doctoral dissertation, 1979.

SEDL (Southwest Educational Development Laboratory). "Planning Information for Follow Through Experiments (RFP No. 78-101) Task 1B-Final." Austin, Tex.: Author, November 19, 1979.

Segalowitz, N. "Psychological Perspectives on Bilingual Education." In Spolsky, B., and Cooper, R.L. (eds.), Current Trends in Bilingual Education, 1975.

Serdyuchenko, G.P. "The Eradication of Illiteracy and the Creation of New Written Language in the USSR." International Journal of Adult and Youth Education, 1962, 14, 23-29.

Shepard, M., and Baker, K. Year Round Schools. Lexington, Mass.: D.C. Heath, 1977.

Skoczylas, Rudolph V. "An Evaluation of Some Cognitive and Affective Aspects of a Spanish-English Bilingual Education Program." Unpublished doctoral dissertation, University of New Mexico, 1972.

Skutnabb-Kangas, Tove. Language in the Process of Cultural Assimilation and Structural Incorporation of Linguistic Minorities. Washington, D.C.: National Clearinghouse for Bilingual Education, 1979.

Smith, Hyram C., and Smith, Van Nguyen. "Evaluating Program Effectiveness: Measuring Input, Process, and Product Variables in New York City Federally Funded High School Bilingual Programs." New York City: New York City Board of Education, 1978.

Smither, Nancy. "Program Description of McAllen, Texas, Immersion Project." 1981. Mimeographed.

Snow, C.E., and Hoefnagel-Hohle, W. "The Critical Period for Language Acquisition: Evidence from Second Language Learning." Child Development, 1978, 49, 1114-28.

Solis, Juan D. Evaluation Instruments for Bilingual Education. Austin, Tex.: Dissemination Center for Bilingual Bicultural Education.

South San Francisco Unified School District. "ESEA Title VII Final Evaluation Report. Spruce School Bilingual Education Program." San Francisco, Calif.: Author, 1979.

Spolsky, B. (ed.), The Language Education of Minority Children. Selected Readings. Rowley, Mass.: Newbury House, 1972.

St. John Valley Bilingual Project. "Five Year Evaluation Report 1975-1980, Frenchville, St. Agatha, Van Buren, Madawska." Madawska, Maine: Author, 1980. Mimeographed.

Stebbins, Linda B.; St. Pierre, Robert G.; Proper, Elizabeth C.; Anderson, Richard B.; and Cerva, Thomas R. Education as Experimentation: A Planned Variation Model Volume IV-A An Evaluation of Follow Through. Cambridge, Mass.: ABT Associates, April 1977.

Stern, Carolyn. "Final Report of the Compton Unified School District's Title VII Bilingual-Bicultural Project: September 1969 through June 1975." Compton City, Calif.: Compton City Schools, 1975. Mimeographed.

Stern, H.H. "Bilingual Schools and Foreign Language Education: Some Implications of Canadian Experiments in French Immersion." In Alatis, J.E. (ed.), International Dimensions of Bilingual Education. Washington, D.C.: Georgetown University Press, 1978.

Swain, M. "Bilingual Education for the English-Speaking Canadian." In Alatis, J.E. (ed.), International Dimensions of Bilingual Education. Washington, D.C.: Georgetown University Press, 1978.

Swain, Merrill. "Bilingual Education: Research and Its Implications." Paper presented at the 13th Annual TESOL Convention, 1979.

____________. "French Immersion: Early, Late, or Partial?" Canadian Modern Language Review, 1978, 34, 577-86.

____________. "School Reform Through Bilingual Education: Problems and Some Solutions in Evaluating Programs." Comparative Education Review, 1978, 22, 420-33.

Taylor, Thomasine H. "A Comparative Study of the Effects of Oral-Aural Language Training on Gains in English Language for Fourth and Fifth Grade Disadvantaged." Unpublished doctoral dissertation, University of Texas at Austin, 1969.

Theverin, Tania. "Pedagogical Implications of Language Policy in African Schools: A Review of the Francophone Literature." Washington, D.C.: World Bank, 1980. Mimeographed.

Tikunoff, William J. Second Quarter Report for Significant Bilingual Instructional Features Study. San Francisco: Far West Laboratory for Educational Research and Development, April 1981.

Tikunoff, William J. et al. Review of the Literature for a Descriptive Study of Significant Bilingual Instructional Features. San Francisco, Calif.: Far West Laboratory for Educational Research and Development, March 1980.

Toukomaa, Pertti, and Skutnabb-Kangas, Tove. *The Intensive Teaching of the Mother Tongue to Migrant Children at Preschool Age*. Research Report 26. Tampere, Finland: Department of Sociology and Social Psychology, University of Tampere, 1977.

Trevino, Bertha. "An Analysis of the Effectiveness of a Bilingual Program in the Teaching of Mathematics in the Primary Grades." Unpublished doctoral dissertation, University of Texas at Austin, 1968.

Trevino, Bertha G. "Bilingual Instruction in the Primary Grades." *Modern Language Journal*, 1970, 54, 255-56.

Troike, Rudolph. Personal Communication (via telephone), 1981.

______________ . "Research Findings Demonstrate the Effectiveness of Bilingual Education." *Institute for Cultural Pluralism Newsletter*, September-October 1979, 1(4).

Troike, Rudolph C. *Research Evidence for the Effectiveness of Bilingual Education*. Washington, D.C.: National Clearinghouse for Bilingual Education, 1978.

________________ . "Synthesis of Research on Bilingual Education." *Educational Leadership*, March 1981, 498-504.

Tucker, G.R. "Implications for U.S. Bilingual Education: Evidence from Canadian Research." *Focus*, 1980, 2.

Tucker, G.R. Personal Communication, 1980.

U.S. Commission on Civil Rights. *A Better Chance to Learn: Bilingual/Bicultural Education*. Publication 51. Washington, D.C.: Author, 1975.

U.S. Department of Education. "Educational Services for Students with Limited English Proficiency. A Service Delivery Assessment." Washington, D.C.: Office of Evaluation and Program Management, Division of Program Assessment, 1980. Mimeographed.

__________________________ . "An Evaluation of Project Information Packages (PIPs) as Used for the Diffusion of Bilingual Projects." Executive Summary. Washington, D.C.: Office of Program Evaluation, 1980. Mimeographed.

__________________________ . "Proposed Strategy for the Evaluation of ESEA Title VII Bilingual Education Programs." Washington, D.C.: Office of Bilingual Education and Minority Language Affairs, 1980. Mimeographed.

U.S. Department of Health, Education, and Welfare. *The Condition of Bilingual Education in the Nation*. Washington, D.C.: Office of Education, 1978.

______________________________. "Office of Education Response to 'A Re-examination of the Impact of ESEA Title VII Spanish/English Bilingual Education Program' Conducted by NIE." November 1979. Mimeographed.

______________________________. Education Division. "Proposed Research Plan for Bilingual Education." July, 1979. Mimeographed.

______________________________. Office of Education. National Bilingual Clearinghouse Conferences. October/November, 1976.

Valencia, A. "Bilingual/Bicultural Education—An Effective Learning Scheme for First Grade and Second Grade Spanish Speaking, English Speaking, and American Indian Children in New Mexico." Albuquerque, N.M.: Southwestern Cooperative Educational Laboratory, 1971.

Valencia, Atilano A. "The Relative Effects of Early Spanish Language Instruction on Spanish and English Linguistic Development. An Evaluation Report on the Pecos Language Arts Program for Western States Small School Projects." Albuquerque, N.Mex.: Southwestern Cooperative Educational Laboratory, 1970.

Van Slambrarck, Paul. "Texas Debates Blueprint for 'Model' Bilingual Classes." The Christian Science Monitor, Thursday, May 28, 1981.

Veilleux, David. "A Survey of Program Impact: Of the St. John Valley Bilingual Education Program, 1970-1977." Modawsta, Maine, 1977.

Veltman, C.J. "Relative Educational Attainment of Hispanic-American Children, 1976." Paper presented at the Aspira Hispanic Forum for Responsive Educational Policy, Washington, D.C., October 1980.

Veltman, Calvin. "The Role of Language Characteristics in the Socioeconomic Attainment Process of Hispanic Origin Men and Women." Washington,D.C.: National Center for Education Statistics, 1980.

Venezky, Richard L. "Non-standard Language and Reading." Elementary English, 1970, 47, 334-45.

Verdoodt, Albert. "Education in a Multilingual and Multicultural Context." Educational Documentation and Information, 1977, 204.

Voiri, Lillian, and Rosier, Paul. "Rock Point Community School: An Example of a Navajo-English Bilingual Elementary School Program." TESOL Quarterly, 1978, 12(3), 263-69.

Von Maltitz, Frances Willard. Living and Learning in Two Languages: Bilingual/Bicultural Education in the United States. New York: McGraw-Hill, 1975.

Vygotsky, L.S. Thought and Language. Cambridge, Mass.: MIT Press, 1962.

Waggoner, Dorothy. "Non-English Language Background Persons: Three U.S. Surveys." In TESOL Quarterly, September 1978, 12(3), 247-62.

Wald, Benji. "Bilingualism." In Siegel, B. (ed.), Annual Review of Anthropology, Vol. III, Palo Alto, Calif.: Annual Reviews, Inc. 1974, 301-321.

Whorf, B.L. Language, Thought and Reality, J.B. Carroll (ed.). Cambridge: MIT Press; and New York: Wiley, 1956.

Wiley, D.E., and Harnischfeger, A. "Explosion of a Myth: Quantity of Schooling and Exposure to Instruction, Major Educational Vehicles." Educational Researcher, 1974, 3, 7-12.

Willink, E.W. "A Comparison of Two Methods of Teaching English to Navajo Children." Unpublished doctoral dissertation, University of Arizona, 1968.

Winter, David L. "A Study Analyzing the Effect a School's Bilingual-Bicultural Program Has on Student Achievement." Unpublished doctoral dissertation, University of Wyoming, 1979.

Wong-Fillmore, Lily. "Second Quarterly Report 1980-81 for Learning English Through Bilingual Instruction." Mimeographed.

Yoloye, E.A. "Evaluation of Six-Year Yoruba Primary Project." Paper presented at the Seminar of Government Representatives of Five Yoruba Speaking States of Nigeria, University of Ife, Nigeria, 1977.

Young, John. "Evaluation Report Submitted to the Eastchester Union Free School District." Eastchester Union Free School District, 1980. Mimeographed.

Zappert, L.T., and Cruz, B.R. Bilingual Education: An Appraisal of Empirical Research. Berkeley, Calif.: The Berkeley Unified School District, 1977.

Zirkel, Perry Alan. "An Evaluation of the Effectiveness of Selected Experimental Bilingual Education Programs in Connecticut." Hartford, Conn.: Department of Education, 1972.